

Département des affaires économiques et sociales  
Division de statistique

Études méthodologiques Série F n° 82/Rev.1

# **Manuel de vérification des recensements de la population et de l'habitation**

## **Révision 1**



Nations Unies  
New York, 2010

## Département des affaires économiques et sociales

Le Département des affaires économiques et sociales du Secrétariat des Nations Unies est une interface vitale entre les politiques mondiales dans les domaines économique, social et environnemental et l'action nationale. Ce département opère dans trois grands domaines interdépendants : i) il recueille, produit et analyse une large gamme de données économiques, sociales et environnementales et d'informations sur lesquelles les États Membres de l'Organisation des Nations Unies se fondent pour étudier des problèmes communs et faire le point des options; ii) il facilite les négociations entre États Membres au sein de nombreux organismes intergouvernementaux sur des mesures conjointes à prendre pour faire face à des difficultés mondiales courantes ou nouvelles; et iii) il conseille les gouvernements intéressés sur les moyens de passer de cadres de politiques conçus lors des conférences et des sommets des Nations Unies à des programmes au niveau des pays et, par le biais de l'assistance technique, il aide au renforcement des capacités nationales.

### Note

Les appellations utilisées et la présentation des données correspondantes n'impliquent de la part du Secrétariat de l'Organisation des Nations Unies aucune prise de position quant au statut juridique des pays, des territoires, villes ou zones, ou de leurs autorités, ni quant au tracé de leurs frontières ou limites.

Le terme « pays » utilisé dans la présente publication s'entend également, suivant le cas, des territoires ou zones.

Les appellations « régions développées » et « régions en développement » sont employées à des fins statistiques et n'expriment pas nécessairement une opinion quant au stade de développement de tel ou tel pays ou de telle ou telle région.

Les cotes de documents de l'Organisation des Nations Unies se composent de lettres majuscules et de chiffres.

ST/ESA/STAT/SER.F/82/Rev.1

PUBLICATION DES NATIONS UNIES

Numéro de vente : F.09.XVII.11

ISBN 978-92-1-261224-9

Copyright © Nations Unies, 2010

Tous droits réservés

Imprimé à l'Organisation des Nations Unies, New York

## Préface

L'Organisation des Nations Unies élabore depuis plusieurs années une série de manuels et de rapports techniques destinés à aider les pays à concevoir et à réaliser des recensements de la population et de l'habitation en améliorant leur qualité et leur efficacité. Ces manuels et ces rapports ont été révisés à plusieurs reprises et mis à jour de façon à tenir compte des nouvelles évolutions et des problèmes se faisant jour en matière de recensement. La présente publication fait partie d'une série de manuels qui ont été mis au point afin d'aider les pays à préparer le recensement de l'an 2000 et les recensements futurs. Parmi les autres manuels de la série figurent :

- a) *Handbook on Census Management for Population and Housing Censuses*, Series F, n° 83 (publication des Nations Unies, numéro de vente : E.00.XVII.15);
- b) *Handbook on Geographic Information Systems and Digital Mapping*, Series F, n° 79 (publication des Nations Unies, numéro de vente : E.00.XVII.12);
- c) *Guide for the Collection of Economic Characteristics in Population Censuses* (à paraître).

La publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitation, deuxième révision* (Nations Unies, 2008) examine les avantages d'un système de contrôle et d'amélioration de la qualité, mis en place dès le début d'un processus de recensement, et dont l'importance est décisive pour le succès de l'ensemble des opérations. Elle étudie en outre le rôle d'un plan de vérification qu'il convient d'élaborer dans le cadre du programme de recensement et d'intégrer aux différents plans et procédures mis en œuvre à cet effet. Les utilisateurs du *Manuel de vérification des recensements de la population et de l'habitation* trouveront particulièrement utile de se référer à la publication *Principes et recommandations* qui fournit une grande quantité de renseignements d'ordre général concernant les procédures de vérification présentées aux chapitres III, IV et V.

La présente publication a pour objet de donner aux pays un vaste aperçu de la méthodologie de vérification des données des recensements et des enquêtes et de fournir aux responsables concernés des indications quant à la mise en œuvre des différentes approches en matière de vérification des recensements. Un autre but consiste par ailleurs à inciter les pays à conserver un historique de leurs expériences de vérification, à promouvoir les échanges de vues entre spécialistes du sujet et informaticiens et à documenter les activités menées dans le cadre du recensement ou de l'enquête du moment afin d'éviter la répétition inutile des mêmes tâches lors de la prochaine enquête ou du prochain recensement.

Le *Manuel* passe en revue les avantages et les inconvénients des vérifications manuelles et informatisées. Pour les grands recensements, la correction manuelle des données est, d'un point de vue économique, rarement envisageable. Les conditions d'exécution de ces corrections sont généralement spécifiées dans des logiciels spécialisés qui décèlent et corrigent automatiquement les erreurs et imputent des valeurs calculées à partir d'autres renseignements sur la personne ou le ménage en cause ou bien

sur d'autres personnes ou d'autres ménages. La plus grande partie du *Manuel* traite de la correction automatique des données.

Les vérifications informatisées jouent un rôle majeur en matière de détection et de correction des erreurs. Au stade des vérifications informatisées, des contrôles de cohérence détaillés peuvent être définis en se concertant avec les spécialistes du sujet. Les erreurs décelées peuvent être corrigées soit en se référant au questionnaire d'origine, soit de façon automatique. Bien que les vérifications automatisées accélèrent le traitement, elles exigent un contrôle rigoureux de la qualité des données d'entrée.

La présente publication comprend une introduction et cinq chapitres. L'introduction décrit le processus du recensement, ainsi que les différents types d'erreurs observées dans ce cadre. Le chapitre I traite des règles fondamentales de vérification des recensements. Les chapitres II à V décrivent les procédures et techniques de vérification des données de recensement, aux différents stades de leur traitement. Enfin, des considérations techniques, concernant notamment la programmation, figurent dans les annexes.

Bien que ce manuel soit axé sur la vérification des recensements de la population et de l'habitation, les concepts et les techniques dont il est question sont pour la plupart également applicables aux enquêtes.

Nous tenons tout particulièrement à remercier le Harvard Center for Population and Development Studies pour le concours apporté par M. Michael J. Levin à l'élaboration de ce manuel. Nous remercions également MM. Michael Bankier, Wesley Benjamin, Marcel Bureau, Sean Crowe, Sylvain Delisle et Darryl Janes de Statistique Canada, qui ont revu le projet de publication et contribué par leurs précieux commentaires à sa mise au point finale.

# Table des matières

Préface .....	iii
Introduction .....	1
A. Objet du <i>Manuel</i> .....	1
B. Le processus de recensement .....	2
C. Erreurs affectant les opérations de recensement .....	3
1. Erreurs de couverture .....	3
2. Erreurs matérielles .....	3
a) Erreurs dues à la conception du questionnaire .....	4
b) Erreurs dues à l'agent recenseur .....	4
c) Erreurs dues aux personnes interrogées .....	4
d) Erreurs de codage .....	5
e) Erreurs de saisie des données .....	5
f) Erreurs dues aux vérifications automatiques .....	5
g) Erreurs d'exploitation des données .....	5
D. Structure du <i>Manuel</i> .....	6
I. Opérations de vérification concernant les recensements et les enquêtes .....	7
A. Rétrospective historique .....	7
B. L'équipe de vérification .....	8
C. Méthodes de vérification : données vérifiées et non vérifiées .....	10
D. Principes de base de vérification .....	12
1. Comment la survérification est-elle préjudiciable .....	15
a) Respect des délais .....	15
b) Charges financières .....	15
c) Distorsion des valeurs vraies .....	16
d) Un sentiment erroné de sécurité .....	16
2. Traitement des données inconnues .....	16
3. Modifications parasites .....	17
4. Détermination de tolérances .....	17
5. Enseignements du processus de vérification .....	18
6. Assurance qualité .....	18
7. Coût des opérations de vérification .....	18
8. Imputation .....	19
9. Archivage .....	19
II. Mise en œuvre des opérations de vérification .....	21
A. Considérations de codage .....	23
B. Vérification manuelle et vérification automatisée .....	27
C. Principes de correction des données .....	30
D. Contrôle de validité et de cohérence .....	33
1. Méthode de vérification descendante .....	33
2. Méthode de vérification portant sur plusieurs variables .....	34
E. Méthodes de correction et d'imputation des données .....	38
1. Technique d'imputation statique ou <i>cold deck</i> .....	39

2.	Technique d'imputation dynamique ou <i>hot deck</i> .....	39
3.	Problèmes posés par l'imputation dynamique (méthode <i>hot deck</i> ) .	43
	<i>a)</i> Considérations géographiques .....	43
	<i>b)</i> Utilisation de rubriques connexes .....	44
	<i>c)</i> Effet de l'ordre des variables sur les matrices .....	44
	<i>d)</i> Complexité des matrices d'imputation .....	44
	<i>e)</i> Établissement de la matrice d'imputation .....	45
	<i>f)</i> Matrices d'imputation normalisées .....	46
	<i>g)</i> Cas de non-utilisation de la méthode d'imputation dynamique .....	49
	<i>h)</i> Quelle doit être la taille des matrices d'imputation ? .....	51
	<i>i)</i> Problèmes posés par une taille excessive de la matrice d'imputation .....	51
	ii) Connaissance précise des tâches effectuées au moyen de la matrice d'imputation .....	51
	iii) Problèmes posés par une matrice d'imputation de taille insuffisante.....	52
	iv) Rubriques difficiles à corriger au moyen de matrices d'imputation .....	52
4.	Contrôle des matrices d'imputation .....	52
	<i>a)</i> Constitution de la matrice statique initiale .....	52
	<i>b)</i> Messages relatifs aux erreurs.....	53
	<i>c)</i> Établissement à la demande de listes d'erreurs .....	54
	<i>d)</i> Nombre de passages du programme de vérification.....	55
5.	Indicateurs d'imputation .....	56
F.	Autres systèmes de vérification .....	58
III.	VÉRIFICATIONS STRUCTURELLES.....	61
A.	Vérifications géographiques.....	61
	1. Emplacement des locaux d'habitation (localité) [PIH] .....	61
	2. Résidences urbaines et rurales (P1L) .....	62
B.	Contrôles de couverture.....	63
	1. Dénombrement de fait et de droit .....	63
	2. Hiérarchie des ménages et des unités d'habitation.....	64
	3. Fragments de questionnaires.....	64
C.	Structure des enregistrements sur l'habitation.....	64
D.	Correspondance entre les enregistrements relatifs à l'habitation et à la population .....	65
	1. Habitations inoccupées et occupées .....	65
	<i>a)</i> Décision de laisser une unité d'habitation inoccupée .....	65
	<i>b)</i> Réexamen des unités d'habitation à plusieurs reprises afin de compléter les questionnaires.....	66
	<i>c)</i> Substitution d'une autre unité de logement pour les personnes manquantes.....	66
	2. Ménages et unités d'habitation figurant en double.....	66
	3. Ménages et unités d'habitation manquants .....	66
	4. Correspondance entre le nombre d'occupants et la somme des occupants .....	67

a)	Quand le nombre d'occupants est supérieur au décompte des occupants .....	67
b)	Contrôle du nombre de personnes par sexe.....	68
c)	Numérotation séquentielle .....	68
5.	Correspondance entre les occupants et le type d'immeuble/ de ménage .....	68
E.	Enregistrements en double.....	68
F.	Populations spéciales.....	69
1.	Personnes vivant en collectivité.....	69
a)	Cas dans lesquels les enregistrements collectifs constituent un type d'enregistrement différent .....	69
b)	Cas dans lesquels une variable distingue les enregistrements collectifs des autres enregistrements .....	70
c)	Cas dans lesquels le code « type de collectivité » est omis .....	70
d)	Cas dans lesquels le code « type de collectivité » est présent, mais où il existe un lien entre toutes les personnes.....	70
e)	Distinction entre différents types de collectivité.....	71
2.	Groupes difficiles à dénombrer .....	71
a)	Migrants saisonniers .....	71
b)	Personnes sans abri .....	71
c)	Itinérants et personnes vivant dans des endroits difficiles d'accès.....	71
d)	Résidents civils temporairement absents de leur pays.....	71
e)	Civils étrangers qui ne franchissent pas quotidiennement la frontière et se trouvent dans le pays à titre temporaire, y compris les personnes sans papiers ou les passagers de navires se trouvant dans un port au moment du recensement.....	72
f)	Réfugiés.....	72
g)	Personnel militaire, naval et diplomatique et membres de leur famille se trouvant hors du pays et personnel militaire, naval et diplomatique étranger et membres de leur famille se trouvant dans le pays .....	72
h)	Civils étrangers qui franchissent quotidiennement la frontière pour travailler dans le pays .....	73
i)	Résidents civils qui franchissent quotidiennement la frontière pour travailler dans un autre pays .....	73
j)	Membres d'équipage de navires marchands et pêcheurs résident du pays se trouvant en mer au moment du recensement (y compris ceux qui n'ont pas d'autre lieu de résidence que leurs quartiers à bord d'un navire).....	73
G.	Détermination du chef de ménage et du conjoint.....	74
1.	Détermination de la variable chef de ménage.....	74
a)	Ordre des liens avec le chef de ménage .....	74
b)	Cas dans lesquels le chef de ménage n'est pas la première personne.....	74
i)	Attribution d'un pointeur à l'enregistrement du chef de ménage .....	75
ii)	Désignation du chef de ménage en tant que personne n° 1 .....	76

iii)	Réattribution de codes de lien avec le chef de ménage de façon à placer celui-ci en première position.....	76
c)	Plusieurs chefs de ménage.....	77
d)	Absence de chef de ménage .....	77
2.	Vérification concernant le conjoint .....	77
a)	Cas d'un conjoint unique dans les sociétés monogames .....	77
b)	Cas de plusieurs conjoints dans les sociétés monogames .....	77
c)	Conjoints dans les sociétés polygames.....	78
d)	Autres caractéristiques des chefs de ménage et des conjoints ...	78
H.	Âge et date de naissance .....	78
1.	Indication de la date de naissance, mais omission de l'âge .....	78
2.	Défaut de concordance de l'âge et de la date de naissance .....	78
I.	Décompte des données invalides .....	78
IV.	Vérifications concernant les données sur la population .....	81
A.	Caractéristiques démographiques .....	83
1.	Lien avec le chef de ménage.....	83
a)	Vérifications du lien avec le chef de ménage.....	83
b)	Cas dans lesquels le chef de ménage doit figurer en premier....	84
c)	Cas dans lesquels les liens avec le chef de ménage sont codés « à l'envers » .....	84
d)	Présence de plusieurs conjoints (polygamie) .....	84
e)	Présence de plusieurs parents .....	84
f)	Cas dans lesquels les recensements recueillent des données de lien avec le chef de ménage définissant le sexe .....	85
g)	Défaut de concordance entre le lien avec le chef de ménage et l'état matrimonial .....	85
2.	Sexe.....	85
a)	Cas dans lequel le code du sexe est valide, mais où le chef de ménage et le conjoint sont du même sexe.....	86
b)	Cas dans lequel une donnée sur la fécondité est associée à un homme ou ne figure pas en ce qui concerne une femme adulte.....	86
c)	Code du sexe non valide et présence d'un conjoint .....	86
d)	Non-validité du code du sexe du conjoint .....	87
e)	Code du sexe non valide et présence d'indications quant au sexe féminin.....	87
f)	Code du sexe non valide, lorsqu'il s'agit du mari du conjoint..	87
g)	Code du sexe non valide en l'absence d'indications suffisantes pour déterminer le sexe.....	87
h)	Remarques concernant les rapports des variables imputées à la variable sexe .....	87
3.	Date de naissance et âge .....	88
a)	Âge et date de naissance.....	88
b)	Relation entre âge et date de naissance.....	89
c)	Cas dans lequel l'âge calculé dépasse la valeur limite supérieure fixée.....	89
d)	Vérification de l'âge .....	89



e)	Vérification de l'âge en présence du chef de ménage et du conjoint.....	90
f)	Vérification de l'âge du chef de ménage en l'absence du conjoint, mais en présence d'un enfant.....	90
g)	Vérification de l'âge du chef de ménage en présence d'un parent de ce dernier.....	91
h)	Vérification de l'âge du chef de ménage en présence d'un petit-enfant de ce dernier.....	91
i)	Vérification de l'âge du chef de ménage en l'absence de toute autre donnée d'âge.....	91
j)	Vérification de l'âge du conjoint après détermination de l'âge du chef de ménage.....	92
k)	Vérification de l'âge des autres couples mariés du ménage, lorsque l'âge de l'une des personnes est connu.....	92
l)	Vérification de l'âge d'un enfant après détermination de l'âge du chef de ménage.....	92
m)	Vérification de l'âge d'un parent après détermination de l'âge du chef de ménage.....	93
n)	Vérification de l'âge d'un petit-enfant après détermination de l'âge du chef de ménage.....	93
o)	Vérification de l'âge de toutes les autres personnes.....	93
4.	État matrimonial.....	93
a)	Vérification de l'état matrimonial.....	94
b)	Attribution de la variable état matrimonial sans imputation dynamique.....	94
c)	Attribution de la variable état matrimonial avec imputation dynamique.....	94
d)	Le conjoint devrait être marié.....	94
e)	Conjoint dans un couple marié.....	94
f)	S'il a la qualité de conjoint, le chef de ménage doit être marié.....	95
g)	Chef de ménage, sans conjoint ni enfants.....	95
h)	Imputation en cas d'échec de toutes les autres tentatives.....	95
i)	Relation entre l'âge et l'état matrimonial des jeunes.....	95
5.	Âge au premier mariage.....	95
a)	L'âge au mariage des personnes qui n'ont jamais été mariées doit être laissé en blanc.....	96
b)	Une valeur doit être indiquée dans le cas des personnes qui ont déjà été mariées.....	96
6.	Fécondité : nombre d'enfants et nombre d'enfants vivants.....	96
a)	Informations recueillies sur la fécondité.....	97
b)	Principes généraux applicables à la vérification des données sur la fécondité.....	97
c)	Relation entre le nombre total d'enfants et le nombre d'enfants vivants.....	98
d)	Vérification lorsque seul le nombre d'enfants est indiqué.....	98
e)	Vérification en cas d'indication du nombre d'enfants et du nombre d'enfants vivants.....	99
f)	Vérification en cas de mention du nombre total d'enfants, du nombre d'enfants vivants et du nombre d'enfants décédés..	101

i)	Indication des trois informations.....	101
ii)	Indication de deux informations .....	101
iii)	Indication d'une information .....	101
iv)	Absence de toute information .....	102
g)	Vérification en cas d'indication du nombre d'enfants, du nombre d'enfants vivant à la maison, du nombre d'enfants vivant à l'extérieur et du nombre d'enfants décédés.....	102
i)	Indication des quatre informations.....	102
ii)	Indication de trois des quatre informations .....	103
iii)	Indication de deux informations sur quatre .....	103
iv)	Indication d'une seule information.....	105
v)	Indication d'aucune information.....	105
h)	Cas particulier d'indication de cinq éléments d'informations ou plus .....	106
i)	Importance de l'utilisation d'un donneur unique pour toutes les données sur la fécondité.....	106
j)	Relation entre les enfants du ménage, les enfants vivant à la maison et les enfants vivants .....	106
7.	Fécondité : date de naissance du dernier enfant né vivant et naissances survenues dans les 12 mois ayant précédé le recensement.....	106
8.	Fécondité : âge à la première naissance.....	108
9.	Mortalité.....	109
a)	Âge et sexe de la personne décédée.....	109
b)	Cause du décès.....	109
c)	Mortalité maternelle.....	110
d)	Mortalité infantile .....	110
10.	Orphelins de mère ou de père (P5G) et numéro de ligne de la mère	110
B.	Caractéristiques des migrations .....	111
1.	Lieu de naissance.....	112
a)	Liens entre les données concernant le lieu de naissance et le nombre d'années passées dans la circonscription administrative.....	112
b)	Attribution de la valeur « inconnu » aux données non valides concernant le lieu de naissance .....	112
c)	Utilisation de la méthode d'imputation statique pour le lieu de naissance.....	112
d)	Utilisation de la méthode d'imputation dynamique pour le lieu de naissance .....	113
e)	Attribution du lieu de naissance lorsque la mère d'une personne est présente.....	113
f)	Attribution du lieu de naissance dans le cas d'un enfant du chef de ménage.....	113
g)	Attribution d'un lieu de naissance à un enfant autre qu'un enfant du chef de ménage.....	114
h)	Attribution d'un lieu de naissance aux femmes adultes mariées.....	114

i)	Attribution d'un lieu de naissance aux femmes adultes non mariées.....	114
j)	Attribution d'un lieu de naissance à des individus de sexe masculin.....	114
2.	Nationalité.....	115
a)	Vérification de la nationalité.....	115
b)	Relation entre groupe ethnique/race et nationalité.....	115
c)	Relation entre naturalisation et nationalité.....	115
d)	Rapport entre durée de résidence et nationalité.....	116
3.	Durée de résidence.....	116
a)	Vérification de la durée de résidence.....	116
b)	Résidence de fait/de droit et durée.....	116
c)	Relation entre l'âge et la durée de résidence.....	116
d)	Relation entre lieu de naissance et durée de résidence.....	117
e)	Pour les personnes qui ont toujours vécu au même endroit.....	117
f)	Durée de résidence d'une personne d'après celle de la mère.....	117
g)	Durée de résidence d'une personne d'après celle d'un enfant ...	118
h)	Durée de résidence d'une personne en l'absence de toute autre information disponible.....	118
4.	Lieu de résidence antérieure.....	118
a)	Vérification du lieu de résidence antérieure.....	119
b)	Lieu de résidence antérieure lorsque les frontières ont changé..	
c)	Cas où la personne n'a pas déménagé depuis sa naissance.....	119
d)	Utilisation d'autres personnes faisant partie de l'unité d'habitation.....	119
e)	Aucune autre personne appropriée pour la vérification du lieu de résidence antérieure.....	119
5.	Lieu de résidence à une date déterminée du passé.....	120
6.	Année d'arrivée.....	120
a)	Relation entre l'âge et l'année d'arrivée.....	120
b)	Relation entre le lieu de naissance et l'année d'arrivée.....	121
c)	Cas des personnes qui ont toujours vécu au même endroit.....	121
d)	Détermination de l'année d'arrivée d'une personne à partir de l'année d'arrivée de sa mère.....	121
e)	Détermination de l'arrivée d'un enfant à partir de l'année d'arrivée du chef de ménage.....	122
f)	Détermination de l'arrivée d'une personne en l'absence de toute autre information disponible.....	122
7.	Relation entre la durée de résidence et l'année d'arrivée.....	123
8.	Résidence habituelle.....	123
C.	Caractéristiques sociales.....	124
1.	Aptitude à lire et à écrire (alphabétisation) [P6A].....	124
2.	Fréquentation scolaire (P6B).....	125
a)	Vérification de la fréquentation scolaire.....	125
b)	Fréquentation scolaire à temps plein ou à temps partiel.....	126
c)	Cohérence entre la fréquentation scolaire et l'activité économique.....	126

	<i>d)</i> Attribution d'une valeur en cas de réponse non valide ou incohérente concernant la « fréquentation scolaire » .....	126
3.	Niveau d'instruction (nombre d'années d'étude ou niveau d'étude atteint) .....	126
	<i>a)</i> Vérification du niveau d'instruction.....	126
	<i>b)</i> Âge minimal compatible avec le niveau d'instruction .....	127
	<i>c)</i> Relation entre l'âge et le niveau d'instruction.....	127
4.	Domaine d'étude et diplôme de fin d'études.....	127
5.	Religion .....	127
	<i>a)</i> Vérification concernant la religion .....	128
	<i>b)</i> Indication de religion omise pour le chef de ménage, mais présente pour une autre personne de l'unité d'habitation. ....	128
	<i>c)</i> Indication de religion omise pour le chef de ménage ou pour toute autre personne de l'unité d'habitation .....	128
	<i>d)</i> Personne autre que le chef de ménage, sans indication de religion .....	128
6.	Langues.....	128
	<i>a)</i> Vérification de la langue .....	129
	<i>b)</i> Vérification de la langue : chef de ménage.....	129
	<i>c)</i> Vérification de la langue : personnes autres que le chef de ménage .....	129
	<i>d)</i> Vérification de la langue : utilisation de l'origine ethnique ou du lieu de naissance .....	129
	<i>e)</i> Vérification de la langue : langue maternelle .....	129
	<i>f)</i> Vérification de la langue : aptitude à parler une langue déterminée .....	129
7.	Appartenance ethnique et peuples autochtones .....	130
	<i>a)</i> Vérification concernant l'appartenance ethnique.....	130
	<i>b)</i> Vérification concernant l'appartenance ethnique du chef de ménage.....	131
	<i>c)</i> Vérification concernant l'appartenance ethnique des personnes autres que le chef de ménage .....	131
	<i>d)</i> Vérification concernant l'appartenance ethnique : utilisation de la langue et du lieu de naissance.....	131
8.	Incapacité.....	131
	<i>a)</i> Questions posées dans le cadre du recensement concernant l'incapacité .....	132
	<i>b)</i> Vérification concernant les incapacités .....	132
	<i>c)</i> Incapacités multiples .....	132
	<i>d)</i> Vérification concernant les causes de l'incapacité .....	132
D.	Caractéristiques économiques .....	133
1.	Situation concernant l'activité économique.....	133
	<i>a)</i> Catégories liées à la situation concernant l'activité économique	134
	<i>i)</i> Population au chômage.....	134
	<i>ii)</i> À la recherche d'un emploi.....	135
	<i>iii)</i> Actuellement inactive.....	135
	<i>iv)</i> Causes de la non-recherche d'emploi.....	135
	<i>b)</i> Vérification de la situation au regard de l'activité économique .	135
	<i>i)</i> Personnes pourvues d'un emploi .....	136

ii) Activité économique des personnes sans emploi.....	136
iii) Activité économique des étudiants et des retraités.....	136
iv) Activité économique non valide et déclaration de variables concernant l'emploi .....	136
v) Activité économique non valide et déclaration de variables concernant le chômage.....	136
vi) Activité économique non valide et aucune déclaration concernant les variables économiques.....	137
2. Temps de travail .....	137
3. Profession.....	137
4. Secteur d'activité économique .....	138
5. Situation au regard de l'emploi .....	138
6. Revenu.....	139
7. Secteur institutionnel .....	140
8. Emploi dans le secteur non structuré .....	140
9. Lieu de travail .....	140
V. Vérifications concernant l'habitation .....	143
A. Sujets de base et sujets supplémentaires.....	145
1. Locaux à usage d'habitation : type (sujet de base).....	145
2. Emplacement des locaux à usage d'habitation (sujet de base) .....	146
3. Locaux à usage d'habitation : état d'occupation (sujet de base).....	147
4. Propriété : type (sujet de base) .....	148
5. Nombre de pièces (sujet de base) .....	148
6. Nombre de chambres à coucher (sujet supplémentaire).....	149
7. Espace plancher utilisable (sujet supplémentaire) .....	149
8. Système d'alimentation en eau (sujet de base) .....	150
9. Eau potable : principale source (sujet de base) .....	151
10. Lieux d'aisances : type (sujet de base).....	151
11. Égouts (sujet de base) .....	151
12. Salles d'eau (sujet de base) .....	152
13. Cuisine : disponibilité (sujet de base).....	153
14. Combustible utilisé pour la cuisson des aliments (sujet de base)....	153
15. Éclairage et/ou électricité : type (sujet de base).....	154
16. Évacuation des déchets solides : principal type (sujet de base) .....	154
17. Chauffage : type et énergie utilisée (sujet supplémentaire).....	155
18. Eau chaude : disponibilité (sujet supplémentaire) .....	156
19. Gaz acheminé par conduite : disponibilité (sujet supplémentaire) .	156
20. Utilisation d'une unité d'habitation (sujet supplémentaire).....	156
21. Occupation par un ou plusieurs ménages (sujet de base) .....	157
22. Nombre d'occupants (sujet de base).....	157
23. Bâtiment : type (sujet de base).....	157
24. Année ou période de construction (sujet supplémentaire) .....	158
25. Immeuble : nombre de logements (sujet supplémentaire).....	158
26. Matériaux de construction des murs extérieurs (sujet de base).....	158
27. Matériaux de construction des sols, de la toiture (sujet supplémentaire) .....	159
28. Ascenseur : disponibilité (sujet supplémentaire).....	159
29. Bâtiment à usage agricole (sujet supplémentaire) .....	160

30.	État d'entretien (sujet supplémentaire) .....	160
31.	Caractéristiques du chef de ménage ou d'un autre membre du ménage utilisé comme référence (sujet de base).....	161
32.	Modalités de jouissance (sujet de base) .....	161
33.	Loyer et coût de logement des propriétaires occupants (sujet supplémentaire) .....	161
34.	Unités d'habitation meublées ou non meublées (sujet supplémentaire) .....	162
35.	Dispositifs d'application des technologies de l'information et de la communication : disponibilité (sujet de base) .....	162
36.	Automobiles : nombre (sujet supplémentaire).....	163
37.	Équipements ménagers : disponibilité (sujet supplémentaire).....	163
38.	Espaces extérieurs : disponibilité (sujet supplémentaire).....	164
B.	Unités d'habitation occupées et inoccupées .....	164

### Annexes

I.	Variables calculées.....	165
A.	Variables calculées associées aux données sur l'habitation.....	165
1.	Revenu du ménage .....	165
2.	Revenu familial .....	166
3.	Noyau familial .....	166
4.	Type de famille.....	167
5.	Composition des ménages .....	167
6.	Composition de la famille .....	168
7.	Ménage et état matrimonial.....	168
8.	Impact du VIH/sida sur la structure du ménage.....	169
9.	Personnes apparentées .....	170
10.	Travailleurs de la famille .....	170
11.	Installations complètes de plomberie .....	170
12.	Installations de cuisines complètes .....	171
13.	Loyer brut .....	171
14.	Indice de richesse .....	171
B.	Variables calculées associées aux données sur la population.....	172
1.	Situation de l'activité économique .....	172
2.	Propres enfants.....	172
3.	Parents vivant à la maison.....	173
4.	Scolarisation durant l'année en cours.....	173
5.	Mois écoulés depuis la dernière naissance.....	173
II.	Lien entre la présentation du questionnaire et l'introduction des données....	175
III.	Lecture optique ou saisie manuelle des données.....	179
A.	Introduction des données.....	179
1.	Lecture optique .....	179
2.	Saisie manuelle continue .....	180
a)	Introduction continue sans sauts de champ .....	180
b)	Introduction continue avec sauts de champ .....	181
3.	Saisie manuelle discontinue .....	181
B.	Contrôle .....	182
1.	Contrôle par comparaison .....	183

2. Contrôle indépendant .....	184
C. Considérations relatives à la vérification des données saisies par lecture optique.....	184
D. Conclusions.....	185
IV. Exemple d'organigrammes .....	187
V. Méthodes d'imputation.....	193
VI. Progiciels de vérification .....	197
Glossaire .....	201
Références .....	209

### Tableaux

1. Répartition d'un échantillon de population par groupe d'âge de 15 ans et par sexe, avant et après vérification des données.....	10
2. Population et évolution de la population par groupe d'âge de 15 ans, avec mention de données inconnues : 2000 et 2010.....	12
3. Population et évolution de la population par groupe d'âge de 15 ans, sans mention de données inconnues : 2000 et 2010 .....	12

### Figures

1. Exemples de codes communs pour certaines rubriques.....	26
2. Exemple de ménage fictif, avec indication du lien avec le chef de ménage, du sexe et de la fécondité de ses membres.....	30
3. Exemple de ménage : chef de ménage et conjoint du même sexe .....	31
4. Exemple de ménage avec indication des âges de certains membres.....	32
5. Exemple de ménage présentant des incohérences potentielles des âges déclarés .	32
6. Exemple de règles de vérification à variables multiples de certaines caractéristiques démographiques .....	35
7. Exemple comportant un chef de ménage et un conjoint de même sexe dans un jeu de données non vérifiées.....	36
8. Exemple de spécifications de vérification visant à corriger les valeurs indiquées pour la variable sexe, en langage symbolique .....	37
9. Exemple d'analyses de vérification à plusieurs variables, dans le cas d'une veuve très jeune avec trois enfants.....	37
10. Exemple de ménage utilisé pour fournir les données d'entrée d'une imputation dynamique.....	41
11. Matrice initiale d'imputation statique de l'âge d'après le sexe et le lien avec le chef de ménage.....	42
12. Exemple de matrice d'imputation dynamique à la suite d'une modification .....	42
13. Exemple de matrice d'imputation dynamique à la suite de plusieurs modifications.....	43
14. Exemple de défaut d'indication de la langue du chef de ménage et du père du chef de ménage.....	46
15. Valeurs initiales d'une matrice d'imputation dynamique de la langue .....	47
16. Exemple de membres d'un ménage dans lequel aucune langue n'a été attribuée	47
17. Exemple de données concernant un chef de ménage et un enfant avec omission de l'âge de l'enfant .....	49
18. Exemple de données concernant un chef de ménage et un enfant avec omission de l'âge et du niveau de scolarité de l'enfant .....	50

19. Exemple de jeu de valeurs utilisées dans un tableau d'imputation statique et codes imputés correspondants .....	53
20. Exemple de rapport indiquant le nombre d'imputations par erreur .....	53
21. Exemple de rapport d'erreurs contenues dans un questionnaire.....	54
22. Exemple de liste imprimée d'erreurs supplémentaires par questionnaire portant sur plusieurs variables .....	54
23. Exemple de données sur la population comportant des indicateurs associés aux valeurs imputées .....	57
24. Exemple d'indicateurs correspondant à une jeune femme dont l'enregistrement comporte des données sur la fécondité laissées en blanc, avec ajout d'un indicateur .....	58
25. Exemple de ménage dont le chef figure en premier sur la liste .....	74
26. Exemple de ménage dont le chef est classé en cinquième position .....	75
27. Exemple de ménage comportant des données sur la fécondité .....	99
28. Valeurs initiales pour la détermination du nombre d'enfants vivants avec des données valides d'âge et de nombre d'enfants vivants.....	100
29. Exemple de matrices d'imputation à définir pour des couples de valeurs connues	104

#### *Encadrés*

1. Objectifs des opérations de vérification des recensements.....	9
2. Principes importants de correction des données .....	30
3. Principes directeurs concernant les vérifications structurelles.....	62
4. Vérification et imputation de l'âge .....	89

#### *Figures de l'annexe*

A.II.1. Exemple de questionnaire constitué de pages personnelles .....	175
A.II.2. Exemple de séquence de collecte de l'information sur les pages personnelles d'un questionnaire.....	175
A.II.3. Exemple de questionnaire, page concernant un ménage, sur laquelle figurent toutes les personnes composant le ménage.....	176
A.II.4. Exemple de séquence de traitement des informations inscrites sur des pages de questionnaire concernant les ménages, avec plusieurs personnes par page .....	177
A.II.5. Exemple de page concernant un ménage à plusieurs personnes, ne posant aucun problème de saisie .....	177
A.II.6. Exemple de page concernant un ménage à plusieurs personnes, susceptible de poser des problèmes de saisie.....	178
A.IV.1. Exemple d'organigramme de détermination du chef de ménage.....	188
A.IV.2. Exemple d'organigramme de détermination de la présence d'un conjoint dans le ménage .....	189
A.IV.3. Exemple d'organigramme de vérification de la variable sexe concernant le chef de ménage et le conjoint.....	190



# Abréviations

ED	Enfants décédés
EV	Enfants vivants
EVE	Enfants vivant à l'extérieur
EVF	Enfants vivant au foyer
ICIDH	Classification internationale des handicaps, déficiences, incapacités et désavantages
NMI	Nouvelle méthodologie d'imputation
NTE	Nombre total d'enfants
OIT	Organisation internationale du Travail
OMR	Lecture optique de marques
PES	Enquête postdénombrement
ROC	Reconnaissance optique des caractères
SAS	Système d'analyse statistique
SCN	Système de comptabilité nationale
SD	Secteur de dénombrement
SIG	Systèmes d'information géographique
SPSS	Progiciel de statistiques pour les sciences sociales
UNESCO	Organisation des Nations Unies pour l'éducation, la science et la culture



# INTRODUCTION

## A. Objet du *Manuel*

1. Une enquête ou un recensement judicieusement conçu<sup>1</sup>, dont les résultats comportent des erreurs réduites au minimum, est une source d'informations infiniment précieuse pour un pays. Or, l'exactitude des résultats exige l'utilisation de données exemptes dans toute la mesure possible d'erreurs et d'incohérences, en particulier à la suite de la phase de traitement des données.

*L'édition, ou la vérification des recensements de la population et de l'habitation, désigne la procédure qui consiste à détecter erreurs et incohérences dans chaque enregistrement et d'un enregistrement à l'autre, pendant et après la phase de collecte et de saisie, et par ailleurs à ajuster les différents éléments d'information.*

2. Les données des recensements ou des enquêtes ne sont jamais parfaites. Ayant constaté depuis longtemps que les données des recensements et des enquêtes étaient sujettes à caution, les pays ont mis en œuvre différentes procédures pour remédier aux lacunes et aux défauts de cohérence des réponses recueillies. Toutefois, vu l'espacement important des recensements successifs, les procédures utilisées pour vérifier les données font rarement l'objet de descriptions suffisamment détaillées. De ce fait, lors de l'organisation d'un nouveau recensement ou d'une nouvelle enquête, les pays doivent réinventer les méthodes employées à l'occasion d'activités similaires antérieures.

3. Le *Manuel de vérification des recensements de la population et de l'habitation* entend remédier à cette lacune des connaissances dans ce domaine, et fournir aux responsables concernés des indications quant à l'utilisation de différentes méthodes de vérification des recensements; l'objectif poursuivi consiste en outre à encourager les pays à conserver un historique de leurs expériences en la matière, à améliorer la communication entre les spécialistes du sujet et les informaticiens, ainsi qu'à documenter les activités menées à bien dans le cadre des recensements ou des enquêtes en cours afin d'éviter à l'avenir les répétitions inutiles de certaines tâches.

4. Ce manuel est un ouvrage de référence, pour les spécialistes du sujet<sup>2</sup> comme pour les informaticiens ou les méthodologistes, puisqu'ils travaillent en équipes à la définition de règles et de programmes de vérification applicables aux recensements et aux enquêtes. Sa conception s'apparente à celle d'un « livre de recettes », qui permet aux pays d'adopter les vérifications les mieux adaptées à la situation statistique du moment au niveau national. La présente publication est conçue en outre pour améliorer la communication entre spécialistes du sujet et informaticiens au fur et à mesure de l'élaboration et de la mise en place de leur programme de vérification.

<sup>1</sup> Un recensement est un dénombrement complet. Généralement, une enquête dénombre une plus faible proportion de la population totale. Les vérifications décrites ici devraient pouvoir s'appliquer à l'une ou l'autre de ces activités.

<sup>2</sup> Selon la définition retenue dans le présent manuel, les spécialistes du sujet désignent les personnes dont les activités sont axées sur la population, l'habitation et les domaines connexes.

5. L'introduction décrit le processus du recensement, les divers types d'erreurs rencontrés dans ce cadre, et les principes fondamentaux des opérations de vérification. Les chapitres suivants décrivent les procédures et les techniques de vérification des données du recensement aux différentes étapes du traitement. Bien que ce manuel soit axé sur la vérification des recensements de la population et de l'habitation, les concepts et techniques dont il est question sont pour la plupart également applicables aux enquêtes.

## B. Le processus de recensement

6. Un recensement de la population et/ou de l'habitation recouvre l'ensemble des opérations qui consistent à recueillir, à exploiter, à évaluer, à analyser et à diffuser des données démographiques et/ou relatives à l'habitation, économiques et sociales concernant l'ensemble des personnes et des logements (Nations Unies, 2007). Traditionnellement, le recensement est réalisé à un moment donné, sur la totalité ou une partie bien déterminée du territoire national. Récemment, certains pays ont commencé à effectuer des enquêtes continues pour couvrir l'ensemble du pays, utilisant un formulaire « long » pour assurer à terme une couverture complète. Dans un cas comme dans l'autre, le recensement réalise un instantané de la situation de la population ou de l'habitation à un moment donné.

7. Un recensement a essentiellement pour objet de fournir des informations sur l'effectif, la répartition et les caractéristiques de la population d'un pays. Les données du recensement servent à l'élaboration des politiques, à la planification et à l'administration, comme à la gestion et à l'évaluation des programmes dans différents domaines : éducation, main-d'œuvre, planification familiale, logement, santé, transport et développement rural. La délimitation des circonscriptions électorales et la répartition des participants aux organes directeurs en constituent une utilisation administrative fondamentale. Le recensement est en outre une source précieuse d'informations à des fins de recherche : il fournit en effet des données utilisées pour l'analyse scientifique de la composition et de la répartition de la population, comme pour la modélisation statistique à des fins de prévision de sa croissance future. Le recensement fournit par ailleurs à l'industrie et aux entreprises les données de base nécessaires à l'évaluation de la demande de logements, d'écoles, de mobilier, de nourriture, de vêtements, d'équipements de loisir, de fournitures médicales et de différents produits et services.

8. Tous les recensements et toutes les enquêtes ont en commun un certain nombre de caractéristiques importantes notamment en matière de : *a*) travaux préparatoires; *b*) dénombrement ou collecte; *c*) traitement des données, notamment saisie (introduction au clavier ou lecture optique), vérification et mise en tableaux; *d*) établissement de bases de données et diffusion des résultats; *e*) évaluation des résultats; et *f*) analyse des résultats.

9. Le travail de préparation comprend de nombreuses tâches : détermination du fondement juridique du recensement; établissement du budget, élaboration du calendrier; organisation administrative; cartographie; création d'une liste des unités d'habitation; mise au point du programme d'exploitation; préparation du questionnaire et élaboration de plans et formation du personnel, pour le dénombrement, les tests préalables, le traitement des données et la diffusion des résultats.

10. Les opérations de dénombrement dépendent de la méthode choisie à cet effet, de leur calendrier et de leur durée, du niveau de contrôle et de l'utilisation éventuelle d'un échantillon. Une fois les données recueillies, elles doivent être codées, saisies, vérifiées et exploitées. Le traitement des données produit des bases de microdon-

nées et de macrodonnées. Les bureaux nationaux de recensement/statistique utilisent ces bases de données pour réaliser des mises en tableaux, des analyses de séries chronologiques, des graphiques et des cartes; ils font appel par ailleurs aux systèmes d'information géographique (SIG) pour les opérations de cartographie thématique, ainsi qu'à différentes techniques de diffusion. La teneur et la couverture des résultats sont évaluées par différentes méthodes, notamment l'analyse démographique et les études postdénombrement. Enfin, les résultats font l'objet de différents types d'analyses, notamment de résumés descriptifs, d'analyses opérationnelles, et d'études détaillées d'un ou plusieurs aspects de la situation démographique et sociale du pays.

## C. Erreurs affectant les opérations de recensement

11. Les données du recensement sont entachées d'erreurs dont les causes peuvent être multiples et que l'on peut classer généralement en erreurs de couverture et en erreurs matérielles.

### 1. Erreurs de couverture

12. Les erreurs de couverture sont dues à des omissions ou à des doubles comptages de personnes ou d'unités d'habitation lors du dénombrement. Parmi les causes des erreurs de couverture figurent le caractère inexact ou incomplet des cartes ou des listes des secteurs de dénombrement, l'absence d'inspection directe par les agents de recensement de toutes les unités des secteurs qui leur ont été assignés, les doubles comptages, l'omission des personnes qui ne souhaitent pas être recensées, le traitement erroné de certaines catégories de personnes telles que les visiteurs et les étrangers non résidents et la perte ou la destruction d'enregistrements de recensement après dénombrement. Les erreurs de couverture doivent, dans toute la mesure possible, être résolues sur le terrain. Le processus de vérification administrative supprime les données effectivement dédoublées. Toutefois, il faut prendre soin d'établir s'il s'agit d'un double enregistrement de ménages ou de personnes. Les jumeaux, par exemple, peuvent donner lieu à des informations identiques, à l'exception du numéro séquentiel. Par conséquent, les règles de vérification appliquées dans le cadre de ce processus consistent à déterminer quand accepter et quand rejeter des informations qui font apparemment double emploi et quand introduire des modifications par imputation.

13. Les vérifications structurelles décrites au chapitre III consistent à contrôler si les ménages correspondent au nombre adéquat d'enregistrements personnels correctement classés et si des personnes ont pu être comptées deux fois.

### 2. Erreurs matérielles

14. Les erreurs matérielles sont imputables à la notification ou à l'enregistrement inexact des caractéristiques des personnes, des ménages et des logements. La cause de ce type d'erreur peut être un défaut de conception des questions ou le choix d'un ordre séquentiel inapproprié ou encore un défaut de communication entre recensé et recenseur; il peut s'agir également d'erreurs de codage ou de saisie, et d'erreurs commises lors des contrôles manuels et informatisés ou lors de la mise en tableaux des résultats. À chaque étape du processus, il importe donc de réaliser soigneusement et de conserver des vérifications rétrospectives (appelées également analyses rétrospectives) pour veiller à ce qu'il n'y ait aucune perte de données. On trouvera dans les sections suivantes une description de chacune des erreurs mentionnées ci-dessus.

a) Erreurs dues à la conception du questionnaire

15. La mauvaise formulation des questions ou des instructions figure parmi les causes d'erreurs matérielles. Le type de questionnaire, sa présentation et sa formulation précise, ainsi que la disposition des différentes rubriques justifient que le plus grand soin soit apporté à cette tâche, en raison de l'impossibilité de remédier aux défauts de conception, une fois les opérations de recensement commencées ou terminées. Il conviendrait de procéder à des essais préalables pour réduire au minimum les erreurs potentielles dues à un défaut de conception. Si, par exemple, les instructions de saut de champ ne sont pas claires ou ne figurent pas au bon endroit, le recenseur risque de sauter par erreur certaines sections du questionnaire ou d'omettre ainsi de recueillir toutes les informations correspondantes.

b) Erreurs dues à l'agent recenseur

16. Il existe des interactions entre recenseurs et recensés, sauf en cas d'utilisation d'un questionnaire à remplir soi-même. L'agent de recensement peut se tromper en posant les questions, en abrégant ou en modifiant leur formulation ou en s'abstenant de bien expliquer leur signification à la personne interrogée; il risque également d'introduire des erreurs en notant les réponses. La qualité des recenseurs et de leur formation contribue de façon décisive à la qualité des données recueillies; ils doivent être convenablement formés à tous les aspects des procédures de recensement. Il faut par ailleurs les amener à comprendre l'importance de leur tâche et comment le dénombrement proprement dit s'inscrit parmi les différentes étapes du recensement. Qui plus est, puisque les agents recenseurs proviennent d'horizons très différents et que leur niveau d'instruction est plus ou moins élevé, il faut assurer une formation propre à garantir qu'ils savent comment poser les questions pour obtenir une réponse appropriée.

c) Erreurs dues aux personnes interrogées

17. Des erreurs risquent également d'être introduites dans les données lorsque les personnes interrogées ne comprennent pas certaines questions; elles peuvent aussi résulter d'un enregistrement délibérément erroné ou de réponses approximatives (lorsque quelqu'un d'autre que la personne à laquelle les informations se rapportent fournit les réponses au questionnaire). Il est possible d'améliorer la qualité des réponses individuelles en assurant la publicité du recensement, ainsi que la formation des recenseurs de telle sorte qu'ils puissent expliquer l'objet du recensement et justifier les différentes questions. Certains pays utilisent des questionnaires à remplir soi-même pour supprimer les interactions entre recenseur et recensé. Dans le cas des questionnaires à remplir soi-même, les erreurs sont dues au défaut de compréhension des questions et des instructions par les personnes interrogées.

18. C'est au stade du dénombrement qu'il est le plus facile de remédier aux erreurs imputables aux personnes interrogées et à l'agent de recensement, lorsque formulaires, recensés et recenseurs sont encore accessibles. Le personnel d'encadrement doit être en mesure de former les agents de recensement; il doit également pouvoir vérifier régulièrement les données recueillies par ces derniers pendant les opérations de dénombrement, pour assurer qu'aucun biais systématique n'est introduit dans les données. Enfin, les chefs d'équipe doivent corriger sur le terrain les erreurs des agents recenseurs et des personnes interrogées, avant que les questionnaires ne soient envoyés aux bureaux régionaux ou centraux.

#### d) Erreurs de codage

19. Des erreurs peuvent se produire dans le cadre des opérations de codage, la personne affectée à cette tâche étant susceptible de coder l'information de façon erronée. Les erreurs de frappe peuvent être à l'origine d'erreurs lors de l'introduction des données. En règle générale, l'absence de contrôle et de vérification à ce stade retarde la diffusion des données, la détection et la correction des erreurs étant plus délicates par la suite. Les vérifications manuelles interviennent souvent avant ou pendant le codage.

#### e) Erreurs de saisie des données

20. Des contrôles de vraisemblance et certains contrôles de cohérence peuvent être intégrés au logiciel de saisie de données afin d'éviter l'introduction de données incorrectes. Un système intelligent de saisie des données veille à ce que la valeur affectée à chaque champ ou à chaque élément de donnée soit dans le domaine correspondant de valeurs admissibles. Ce système a pour effet de rendre d'autant plus probable l'introduction de données plausibles par l'opérateur de saisie et d'alléger dans une certaine mesure les tâches de vérification aux stades ultérieurs du processus de mise en forme des données. Ces contrôles risquent cependant de ralentir la saisie. Aussi importe-t-il d'évaluer soigneusement l'importance des contrôles de cohérence à ce stade au regard de la nécessité de maintenir une rapidité adéquate. Il faut donc à cet égard définir à priori un compromis, pour que les opérateurs de saisie ne consacrent pas un temps excessif aux tâches de contrôle. La vérification des données introduites au clavier améliore nécessairement leur qualité. Une deuxième introduction au clavier des mêmes informations, souvent par sondage, permet de vérifier les formulaires saisis. Le lien entre la présentation du questionnaire et l'introduction des données est examiné à l'annexe II et la comparaison entre lecture optique et saisie manuelle est traitée à l'annexe III.

#### f) Erreurs dues aux vérifications automatiques

21. Les vérifications représentent une des étapes décisives du traitement des données du recensement. Le processus de vérification modifie ou corrige les données non valides ou présentant un défaut de cohérence, en imputant des données plausibles aux réponses manquantes ou incohérentes. Paradoxalement, l'une ou l'autre de ces opérations risque d'introduire de nouvelles erreurs.

#### g) Erreurs d'exploitation des données

22. Des erreurs peuvent se produire au stade de la mise en tableaux des données, du fait d'erreurs de traitement ou de l'utilisation d'informations « inconnues » (non spécifiées). Il est difficile de procéder alors à des corrections sans introduire de nouvelles erreurs. Enfin, l'absence de vérifications croisées entre les différents tableaux, et les erreurs d'impression entraînent des erreurs au stade de la publication. Au lieu de chercher à corriger les tableaux proprement dits, il importe de mettre à jour le système de traitement pour effectuer des vérifications supplémentaires dès que des défauts de cohérence apparaissent dans les tableaux. En cas de report d'erreurs à tous les stades du processus jusqu'à celui de la publication, celles-ci seront alors manifestes et la valeur des résultats obtenus sera douteuse. Si des « corrections » sont apportées au stade de la mise en tableaux des données, par exemple, si quelques inconnues diverses apparaissent et figurent dans les « totaux » mais pas dans la distribution, les tableaux ne peuvent être reproduits par d'autres analystes et perdent de leur valeur. Il est plus ju-

dicieux de considérer le traitement du recensement comme un système d'information en retour, afin que les changements soient apportés à l'ensemble de données au stade de la vérification plutôt qu'à celui de la mise en tableaux. Avant de publier les tableaux, il est donc essentiel de procéder à un contrôle rigoureux pour garantir que toutes les tâches d'exploitation de données sont définies pour toutes les unités géographiques. Tandis que les contrôles de vraisemblance et de cohérence réalisés au stade de la vérification permettent de corriger la plus grande partie des erreurs, un contrôle général après dépouillement, parfois appelé « macrovérification », joue un rôle essentiel. Il y a lieu de confier à des personnes dûment formées et expérimentées l'examen des divers tableaux obtenus, afin de vérifier si les chiffres figurant dans les différentes cases sont compatibles avec la situation locale connue. Dans un nombre limité de cas, un contrôle rapide des feuilles de recensement peut faire apparaître des erreurs de codage. Le calcul de certains rapports et taux d'accroissement et la comparaison avec les chiffres des recensements précédents ou avec d'autres données numériques publiées dans le cadre d'enquêtes par sondage peuvent également être utiles. Toutefois, la comparaison avec les chiffres fondés sur d'autres enquêtes ou sur des données administratives ne doit être entreprise que si les concepts utilisés sont comparables. Si des erreurs sont constatées dans les tableaux de résultats, il convient de corriger en premier lieu le jeu de microdonnées, notamment pour permettre à d'autres informaticiens du bureau national de recensement/statistique d'établir des tableaux comparables. En outre, puisque les bureaux nationaux de recensement/statistique diffusent parfois des éléments de fichiers de microdonnées à des chercheurs et à d'autres utilisateurs des secteurs public et privé, la reproductibilité des tableaux doit être assurée.

23. Tel qu'indiqué ci-dessus, le recensement se compose d'un certain nombre d'opérations successives interdépendantes, dont chacune peut comporter des erreurs. Il importe de ne pas perdre de vue que les vérifications automatiques font partie d'un système d'information en retour et qu'elles influent non seulement sur les opérations en aval de mise en tableaux, mais aussi en amont sur les opérations de collecte et de traitement sur le terrain. Aussi l'intensification des vérifications sur le terrain est-elle la meilleure façon pour un bureau national de recensement/bureau de statistique d'éviter les problèmes liés aux vérifications automatisées. Le bureau national de recensement/bureau de statistique doit en outre veiller à l'exactitude du codage et de la saisie des données, ainsi qu'à l'obtention en permanence d'informations en retour concernant toutes les opérations, notamment de saisie, de vérification et de dépouillement.

## **D. Structure du Manuel**

24. Le chapitre I est consacré au rôle des opérations de vérification dans le cadre des recensements et des enquêtes. Les autres chapitres traitent de sujets spécifiques. Le chapitre II décrit la mise en œuvre concrète des opérations de vérification et d'imputation, y compris les méthodes d'imputation. Le chapitre III décrit les vérifications de structure, lesquelles concernent simultanément les informations sur l'habitation et sur la population, ainsi que certaines procédures visant à faciliter les autres vérifications, consistant par exemple à déterminer la présence éventuelle d'un seul chef de ménage. Le chapitre IV traite des vérifications touchant la population et le chapitre V des vérifications qui se rapporte à l'habitation. Enfin, une série d'annexes passe en revue des questions particulières liées à la vérification des données des recensements de la population et de l'habitation et aux opérations d'imputation nécessaires.



# Chapitre I

## Opérations de vérification concernant les recensements et les enquêtes

### A. Rétrospective historique

25. Avant l'arrivée des ordinateurs, la plupart des recensements comportaient l'embauche d'un grand nombre d'employés semi-qualifiés chargés de vérifier chaque bulletin. Toutefois, en raison de la complexité des relations entre un nombre même relativement réduit de rubriques, de simples contrôles ne pouvaient laisser escompter la détection de tous les défauts de cohérence possibles. En effet, des employés différents interprétaient vraisemblablement les règles de différentes façons et tel ou tel employé pouvait même ne pas être parfaitement cohérent.

26. L'apparition des ordinateurs a profondément modifié les opérations de vérification des recensements. Les ordinateurs ont permis de déceler des défauts de cohérence bien plus nombreux que les vérifications manuelles. Les spécifications de vérification sont devenues de plus en plus complexes et évoluées. Les imputations automatisées ont alors été possibles et différentes règles ont été définies à cet effet (Nordbotten, 1963; Naus, 1975). Simultanément, le processus opératoire autorisait des contacts de plus en plus nombreux avec les personnes interrogées ou du moins une multiplication des rapprochements avec les questionnaires dûment complétés. De nombreuses équipes de vérificateurs ont commencé à penser que la qualité du recensement était directement fonction de l'ampleur du travail de vérification et que les résultats étaient d'autant plus exacts que ce même travail était plus sophistiqué. Les programmes ont donc généré des milliers de messages d'erreur, exigeant un examen manuel des questionnaires d'origine ou pour certaines enquêtes de nouvelles entrevues avec les personnes interrogées<sup>3</sup>.

27. Grâce aux ordinateurs, il est devenu de plus en plus facile d'introduire des modifications dans le jeu de données. Parfois ces modifications ont eu pour effet de corriger certains enregistrements ou certaines rubriques. Nombre d'enregistrements faisaient l'objet de plusieurs traitements, les erreurs et les défauts de cohérence étant examinés à chaque fois par différentes personnes (Boucher, 1991; Granquist, 1997).

28. Toute cette évolution a conduit à l'élaboration de plusieurs programmes généraux de vérification des recensements, dont certains sont encore utilisés aujourd'hui. À l'origine les programmes ont été conçus pour de gros ordinateurs : certains ont par la suite été modifiés pour être utilisés sur des ordinateurs personnels. Au cours de cette période, Fellegi et Holt (1976) ont mis au point une nouvelle méthode générale de vérification et d'imputation, qui n'a pas été immédiatement appliquée, mais qui est actuellement adoptée de plus en plus fréquemment, à la faveur de la complexification des opérations de vérification effectuées par les bureaux nationaux de recensement/statistique.

<sup>3</sup> Un examen des diverses méthodes d'imputation est présenté à l'annexe V, et des progiciels de vérification informatisée sont examinés à l'annexe VI.

29. Un progrès majeur en matière de vérification des recensements est apparu dans les années 80 quand les bureaux nationaux de recensement/statistique ont commencé à utiliser les ordinateurs personnels pour saisir, vérifier et exploiter leurs données. Rapidement, le personnel chargé du traitement des données a pu effectuer les vérifications en direct, au stade de la saisie des données ou peu de temps après. Pour les enquêtes et les recensements effectués dans de petits pays, les informaticiens ont été en mesure d'élaborer des programmes de détection des erreurs au cours de la collecte des données ou lors de la saisie directe sur ordinateur. Ainsi, l'informatisation des vérifications a permis de développer des contacts continus avec les personnes interrogées de façon à résoudre les problèmes rencontrés à l'occasion des contrôles (Pierzchala, 1995).

30. Au cours des premières années, la réalisation de vérifications de plus en plus sophistiquées et de plus en plus complètes sur les données des recensements et des enquêtes a donné apparemment des résultats très satisfaisants. Les équipes de vérification ont défini des spécifications de vérification de plus en plus élaborées, et les spécialistes du traitement informatique ont passé des mois à mettre au point des organigrammes ou des graphiques de décision et des programmes de calcul. Les progiciels en question ont rarement été évalués par des analystes. Les programmes de vérification avaient apparemment la capacité de remédier à tous les problèmes dus aux phases antérieures de collecte, de codage et de saisie des données. Toutefois, il est également apparu à nombre d'analystes que, dans de nombreux cas, toutes ces opérations supplémentaires de vérification avaient pour conséquence d'altérer les données ou du moins de retarder les résultats, ou encore d'y introduire une distorsion. Parfois, les passages en machine étaient si nombreux pour corriger dans un premier temps un élément d'information, puis un autre, que les résultats étaient loin d'être comparables à ceux obtenus à partir des données initiales non vérifiées.

31. Pour nombre de recensements et pour les grandes enquêtes, ces vérifications aussi poussées entraînaient des retards considérables. Les commis passaient beaucoup de temps à rechercher manuellement des questionnaires, tandis que les informaticiens poursuivaient la mise au point d'applications destinées à examiner un très petit nombre de cas. Or, comme le signale Granquist (1997), nombre d'études ont établi qu'une grande partie de ce travail supplémentaire correspondait « à des améliorations qualitatives marginales, inexistantes ou même négatives; de ce fait, de nombreux types d'erreurs systématiques importantes ne peuvent être mises en évidence par les opérations de vérification ».

32. Alors que les organisations statistiques nationales continuent à préparer des recensements et des enquêtes, la réalisation de vérifications automatiques approfondies est possible et même probable. La question suivante se pose donc aux bureaux nationaux de recensement/statistique : quel est le niveau approprié de vérification automatisée ?

## **B. L'équipe de vérification**

33. Lorsqu'ils préparent un recensement, les bureaux nationaux de statistique doivent envisager d'introduire différents types d'améliorations potentielles dans la qualité de leur travail. Parmi ces dernières figurent la création d'un groupe de vérification. Ainsi, les opérations de vérification doivent être confiées à une équipe de responsables du recensement, de spécialistes du sujet et d'informaticiens. Cette équipe devrait être formée dès le début des préparatifs en vue du recensement, de préférence au cours de la mise au point du questionnaire. L'équipe de vérification joue un rôle im-

portant dès le début et conserve son rôle tout au long du processus de vérification. Le soin apporté à la création de l'équipe ainsi qu'à la définition et à la mise en place des règles de vérification et d'imputation garantit pour le recensement une rapidité et une efficacité accrues.

#### Encadré 1

##### **Objectifs des opérations de vérification des recensements**

La correction des recensements s'attache aux objectifs suivants :

- ☞ Fournir aux utilisateurs des données de recensement de qualité supérieure;
- ☞ Identifier les types et les causes d'erreurs;
- ☞ Fournir des résultats de recensement dûment rectifiés.

34. Les réunions entre les fonctionnaires chargés du recensement et la communauté des utilisateurs en ce qui concerne les tableaux de résultats et les autres produits peuvent apporter des points de vue intéressants quant aux vérifications nécessaires. Fréquemment, les utilisateurs demandent un tableau ou un type de tableau particulier dont la réalisation exige des vérifications supplémentaires afin d'en éliminer les incohérences potentielles. L'équipe de vérification doit donc prévoir de créer ces tableaux pendant la période initiale de vérification, plutôt que de les créer en tant que tableaux spéciaux, une fois terminé le traitement du recensement. Le fait de définir les règles et les programmes informatiques de vérification dans le cadre d'un essai préalable ou d'une sorte de répétition générale permet de tester les logiciels proprement dits et de raccourcir les délais d'exécution des différentes phases du processus de vérification et d'imputation. L'équipe de vérification évalue alors l'incidence de ces différents processus et prend les mesures correctives éventuellement nécessaires.

35. Spécialistes du sujet et informaticiens doivent collaborer à la définition des règles de vérification et d'imputation. L'équipe de vérification élabore un programme de recherche et de correction des erreurs au début des préparatifs du recensement. Elle établit ensuite par écrit des séries de règles de contrôle et de vérification de cohérence.

36. Spécialistes du sujet et informaticiens doivent collaborer non seulement lors de la définition des règles de vérification et d'imputation, mais aussi à tous les stades du recensement ou de l'enquête, notamment au stade de l'analyse. Les vérifications excessives comportent autant de risques que les vérifications insuffisantes et que la présence d'informations non vérifiées ou incohérentes dans le jeu de données. Par conséquent, les deux groupes de spécialistes doivent prendre la responsabilité d'entretenir correctement leurs bases de métadonnées. L'équipe de vérification doit en outre utiliser efficacement les sources administratives et les données d'enquête disponibles pour améliorer la qualité des recensements ou des enquêtes ultérieures.

37. Les échanges entre spécialistes du sujet et informaticiens étaient limités lorsque les bureaux nationaux de recensement/statistique utilisaient de gros ordinateurs; ils le sont restés pendant un certain temps après l'arrivée des micro-ordinateurs, mais les logiciels informatiques sont devenus plus faciles à utiliser et, à présent, nombre de spécialistes du sujet sont effectivement en mesure de mettre au point et d'expérimenter leurs propres programmes de tabulation et leurs propres vérifications. Sans généralement procéder eux-mêmes au traitement proprement dit des données, les spécialistes du sujet sont souvent bien au fait des étapes du traitement des données mis au point par les informaticiens.

## C. Méthodes de vérification : données vérifiées et non vérifiées

38. Les pays procèdent à des vérifications des données de recensement afin d'en améliorer la qualité et la présentation. Dans ce chapitre, le *Manuel* attire l'attention sur un problème qui se pose aux bureaux nationaux de recensement/statistique lors de la publication de données de recensement non vérifiées. Le type de problèmes rencontrés est illustré au moyen d'un jeu de données fictif.

39. Le bureau national de recensement/statistique d'un pays fictif doit relever le défi consistant à vouloir répondre aux besoins de multiples utilisateurs. Certains d'entre eux peuvent souhaiter la prise en compte, à des fins d'analyse ou de recherche, des valeurs dites « inconnues », tandis que d'autres souhaitent le cas échéant des données comportant un bruit (erreur possible) réduit au minimum pour leurs propres besoins de planification ou d'action. Si le bureau national de recensement/statistique diffuse un tableau non corrigé, comme celui de la moitié gauche du tableau 1, alors analystes et responsables devront faire des hypothèses lorsqu'ils utiliseront ces données. Le tableau 1 illustre ce problème uniquement dans le cas d'un petit nombre de personnes. Dans le pays en question, les données enregistrées comportent une omission du sexe<sup>4</sup> pour 23 personnes et de l'âge pour 15 personnes; ces omissions sont probablement imputables à des non-réponses ou à des erreurs d'introduction au clavier; en particulier, il y a eu dans deux cas omission conjointe du sexe et de l'âge.

4 Dans la version anglaise de cette publication, les mots « Sex » et « gender » sont utilisés indifféremment pour désigner le sexe.

Tableau 1  
Répartition d'un échantillon de population par groupe d'âge de 15 ans et par sexe, avant et après vérification des données

Groupe d'âge	Données non vérifiées				Données vérifiées		
	Total	Hommes	Femmes	Non indiqué	Total	Hommes	Femmes
Total	4147	2033	2091	23	4147	2045	2102
Moins de 15 ans	1639	799	825	15	1743	855	888
15 à 29 ans	1256	612	643	1	1217	603	614
30 à 44 ans	727	356	369	2	695	338	357
45 à 59 ans	360	194	166	0	341	182	159
60 à 74 ans	116	54	59	3	114	53	61
75 ans et plus	34	12	22	0	37	14	23
Non indiqué	15	6	7	2			

40. La plupart des utilisateurs prennent généralement leur propre décision quant au traitement des valeurs inconnues. Une solution logique, parfois naïve, consisterait à répartir les valeurs inconnues dans la même proportion que les valeurs connues. Si le bureau national de recensement/statistique décide d'imputer les données inconnues, l'équipe de vérification peut retenir un effectif de 12 hommes et de 11 femmes, soit une répartition dans des proportions sensiblement égales, mais introduisant un biais, puisque le recensement a dénombré plus de femmes que d'hommes. Les résultats seront alors cohérents avec les données corrigées indiquées dans la partie droite du tableau 1.

41. Il existe d'autres options en matière de traitement des données inconnues. Par exemple, l'équipe de vérification peut décider d'introduire des corrections fondées sur la seule répartition par sexe, en ignorant toutes les autres informations dis-

ponibles, telles que la relation entre les conjoints, le fait de savoir si une personne de sexe non spécifié est signalée comme étant la mère d'une autre personne ou si, pour une personne de sexe non spécifié, il y a une indication de nombre d'enfants nés. Une autre stratégie d'imputation consisterait à prendre en compte une ou plusieurs de ces variables.

42. Le bureau national de recensement/statistique pourrait faire appel à une autre solution consistant à décider de fonder l'imputation sur la répartition par âge. Pour l'échantillon de population du tableau 1, l'âge est omis dans 15 cas au total. Ces données pourraient également être réparties dans les mêmes proportions que les valeurs connues, ce qui constitue là encore une stratégie d'imputation logique. Toutefois, l'équipe de vérification pourrait probablement obtenir de meilleurs résultats en considérant d'autres variables et certaines combinaisons, par exemple les âges comparés du mari et de la femme, des parents et des enfants, ou des grands-parents et des petits-enfants, ou encore la présence d'enfants d'âge scolaire, de retraités et de personnes de la population active.

43. Au tableau 1, les données corrigées de la partie droite sont « plus pures », puisque les données inconnues en ont été éliminées (voir données figurant sous la rubrique « données corrigées »). Les données inconnues ont disparu de cette partie du tableau, puisqu'elles ont été réparties par le programme dans des colonnes correspondant à différentes réponses. Toutefois, nombre de démographes et de spécialistes ont traditionnellement exprimé le vœu de voir figurer des données inconnues dans tous les tableaux, comme au tableau 1. Selon eux, cette façon de procéder leur permet d'effectuer différents types d'évaluations sur les chiffres présentés afin de mesurer l'efficacité des méthodes de recensement ou de faciliter la préparation des enquêtes et des recensements futurs. Or, les deux objectifs peuvent être atteints — un tableau corrigé destiné aux utilisateurs proprement dits et un tableau non corrigé présenté pour appréciation — en créant des tableaux avec et sans mention des données inconnues.

44. Il importe que les bureaux de statistique fassent tout leur possible pour maintenir les données initialement recueillies. Un ensemble complet des données initialement saisies devrait être classé, non seulement en tant qu'élément des enregistrements passés, mais aussi à titre de référence si le personnel décide de révérifier une partie quelconque des données établies depuis le début. Toutefois, les valeurs initiales de certaines rubriques essentielles, telles que l'âge, le sexe et la fécondité, devraient être conservées sur chaque enregistrement afin de permettre aux démographes et à d'autres d'analyser les résultats des vérifications.

45. La prise en compte des données inconnues dans les tableaux publiés soulève une autre difficulté; ces données peuvent en effet avoir une incidence quant à l'analyse des tendances. Les nouvelles techniques disponibles facilitent considérablement ce type d'analyse par rapport à la situation passée. Par exemple, le tableau 2 présente une répartition par âge telle qu'elle ressort de deux recensements consécutifs de ce petit pays, passant de 217, soit quelque 6,5 % des réponses apportées en 2000, à seulement 15, soit moins de 1 % des réponses en 2010.

46. À cet égard, le bureau national de recensement/statistique doit déterminer dans quelle mesure des effectifs incohérents de données inconnues affectent le recensement considéré et l'évolution d'un recensement à l'autre. Par exemple, le chiffre de 6,5 % de données inconnues pour le recensement de 2000 rend difficile la comparaison des répartitions en pourcentage par groupe d'âge de 15 ans dans les résultats des deux recensements. Ainsi, le pourcentage de personnes de 15 à 29 ans augment apparemment en passant de 27 % seulement à 30 % au cours de la décennie; or, la répartition des données inconnues a pu affecter les conclusions de l'analyse.

Tableau 2  
Population et évolution de la population par groupe d'âge de 15 ans,  
avec mention des données inconnues : 2000 et 2010

Groupe d'âge	Effectifs		Variation	Variation en pourcentage	Pourcentage	
	2010	2000			2010	2000
Total	4147	3319	828	24,9	100,0	100,0
Moins de 15 ans	1639	1348	291	21,6	39,5	40,6
15 à 29 ans	1256	902	354	39,2	30,3	27,2
30 à 44 ans	727	538	189	35,1	17,5	16,2
45 à 59 ans	360	200	160	80,0	8,7	6,0
60 à 74 ans	116	89	27	30,3	2,8	2,7
75 ans et plus	34	25	9	36,0	0,8	0,8
Non indiqué	15	217	- 202	- 93,1	0,4	6,5

47. Le tableau modifié (tableau 3) est obtenu après répartition des données inconnues, soit proportionnellement, soit par une autre méthode d'imputation. Les variations numériques et en pourcentage apparaissent alors plus nettement, ainsi que la répartition des groupes d'âge dans les deux recensements. Évidemment, pour obtenir des résultats précis et fiables, les équipes de vérification doivent veiller à la cohérence des corrections introduites dans les résultats des deux recensements et/ou des enquêtes, comme à la cohérence interne. La ligne correspondant aux données non indiquées a disparu.

Tableau 3  
Population et évolution de la population par groupe d'âge de 15 ans,  
sans mention des données inconnues : 2000 et 2010

Groupe d'âge	Effectifs		Variation	Variation en pourcentage	Pourcentage	
	2010	2000			2010	2000
Total	4147	3319	828	24,9	100,0	100,0
Moins de 15 ans	1743	1408	335	23,8	42,0	42,4
15 à 29 ans	1217	952	265	27,8	29,3	28,7
30 à 44 ans	695	578	117	20,2	16,8	17,4
45 à 59 ans	341	230	111	48,3	8,2	6,9
60 à 74 ans	114	109	5	4,6	2,7	3,3
75 ans et plus	37	42	- 5	-11,9	0,9	1,3

## D. Principes de base de vérification

48. On entend par vérification l'examen systématique des réponses invalides ou incohérentes suivi de leur correction manuelle ou automatique (au moyen « d'inconnues » ou par imputation dynamique), selon des règles définies au préalable. Certaines opérations de vérification comportent des corrections manuelles introduites manuellement par le bureau. D'autres opérations de vérification comportent des corrections électroniques réalisées au moyen d'ordinateurs. Les publications de recensements risquent de contenir une certaine quantité d'informations dénuées de sens si

les bureaux nationaux de recensement/statistique omettent de vérifier les résultats des recensements ou des enquêtes. La vérification a pour effet de limiter le nombre des estimations faussées, de faciliter le traitement des données et de renforcer la confiance de l'utilisateur à l'égard des résultats. En outre, de l'avis de Pullum, Harpham et Ozsever (1986), « les opérations de vérification ou d'élimination des erreurs ont essentiellement pour objet de déterminer si les différentes réponses sont cohérentes les unes avec les autres et avec le format de base utilisé pour l'enquête ».

49. Les fichiers de données brutes d'un recensement contiennent des erreurs de nature variée. Il existe deux types d'erreurs du point de vue du traitement des données : celles qui risquent de faire obstacle à la poursuite du traitement et celles qui conduisent à des résultats non valides ou incohérents, sans nécessairement interrompre la séquence logique des traitements ultérieurs. Tel qu'indiqué dans *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* (Nations Unies, 2008, par. 1.311), les erreurs du premier type doivent toutes être corrigées et celles du second doivent l'être dans la mesure du possible. La vérification des recensements au stade du dépouillement a donc essentiellement pour objet d'identifier le plus grand nombre possible d'erreurs et de modifier en conséquence le jeu de données, pour assurer la validité et la cohérence des informations recueillies. Néanmoins, les traitements réalisés ne peuvent corriger toutes les erreurs de recensement, notamment les réponses aux questionnaires qui présentent une cohérence interne, mais qui constituent en fait des erreurs de déclaration de la part des personnes interrogées et des erreurs d'enregistrement de la part des agents recenseurs.

50. Les vérifications relèvent généralement de deux catégories : 1) les vérifications fatales, qui détectent des erreurs avec certitude; et 2) les vérifications avec interrogation, qui attirent l'attention sur des données suspectes (Granquist et Kovar, 1997, p. 420). Les erreurs fatales, détectées par les vérifications du même nom, comprennent les entrées non valides ou manquantes, ainsi que les erreurs dues à des incohérences. En revanche, les vérifications avec interrogation détectent les données qui se situent en dehors de limites de vérification essentiellement subjectives, les données dont la valeur est relativement élevée ou faible par rapport aux autres données du même questionnaire et différentes des données douteuses. Pour préserver la fiabilité du recensement et, en particulier lorsque le bureau national de recensement/statistique décide de diffuser des microdonnées, le processus de vérification doit détecter les erreurs fatales et effectuer les corrections correspondantes. Les vérifications avec interrogation se traduisent difficilement par des corrections et comportent moins d'avantages que la détection et la correction des erreurs fatales; enfin, elles augmentent le coût de l'ensemble des opérations de dépouillement.

51. Puisque dans un recensement toutes les données recueillies sont spécifiquement prises en compte en fonction des besoins des planificateurs et des dirigeants, les vérifications avec interrogation doivent impérativement, lors des opérations de vérification et d'imputation, être résolues dans une plus forte proportion par comparaison aux enquêtes. Néanmoins, afin de déterminer les vérifications finales dont un recensement doit faire l'objet, les démographes doivent étudier les vérifications mises au point pour des recensements pilotes et celles élaborées à l'occasion du dépouillement, de manière à s'assurer que chaque vérification envisagée se caractérise par les coûts et les avantages escomptés. Les analyses de ce type font nécessairement partie de l'évaluation du recensement. Comme le font observer Granquist et Kovar (1997, p. 422), les données concernant « les taux de succès, c'est-à-dire la fraction du nombre d'indications qui entraînent des modifications des données de départ, sont rarement mentionnées dans les évaluations ou les études consacrées au processus de vérification ».

52. Une autre série de techniques et d'expressions se rapportent aux microvérifications et aux macrovérifications. Tel qu'indiqué, les vérifications relatives aux recensements et aux enquêtes détectent les erreurs et incohérences à l'intérieur des enregistrements et d'un enregistrement à l'autre. Le présent manuel décrit les opérations de microvérification, c'est-à-dire les moyens utilisés pour garantir la validité et la cohérence des enregistrements individuels et des relations entre les enregistrements relatifs à un ménage. Une autre méthode intitulée macrovérification a pour objet de contrôler des données agrégées et de vérifier qu'elles présentent également un caractère plausible. Selon cette méthode, les données vérifiées sont portées sur les tableaux et comparées aux fréquences et aux tolérances prévues afin d'identifier divers problèmes concernant les données; si des « erreurs » sont constatées, la macrovérification peut apporter une modification globale aux données agrégées, envoyer un enregistrement unitaire en vue d'un nouveau traitement ou ajouter des microvérifications afin de corriger le problème. Par exemple, un pays donné peut avoir un pourcentage très important de personnes dont l'âge n'est pas indiqué. Après imputation de valeurs de l'âge afin d'obtenir un jeu complet, les contrôles effectués au niveau des macrodonnées, au niveau agrégé, permettraient de s'assurer que des omissions sélectives de la part des personnes d'un certain âge n'entraînent par un biais dans les valeurs imputées. Selon les résultats de l'analyse, l'équipe de vérification pourrait convenir de prendre des mesures afin d'atténuer le risque de biais potentiel. Les macrovérifications, de même que les microvérifications, nécessitent une expérimentation minutieuse avant de pouvoir être utilisées.

53. Comme indiqué, les opérations de vérification doivent autant que possible protéger la qualité des données initiales. L'équipe de vérification doit non seulement veiller à la qualité des données et à l'élimination des erreurs qu'elles contiennent, mais doit en outre protéger le travail effectué sur le terrain par l'administration. Les données initiales doivent en effet être protégées à tous les stades du traitement informatique, si l'équipe de vérification décide qu'il faut réexaminer le processus de correction. Les données d'origine sont réexaminées lorsque l'équipe détecte une erreur systématique dans le processus de vérification, ou bien lorsqu'on constate l'absence ou le doublement d'une partie du jeu de données, dont il faut recommencer la constitution et la vérification.

54. Parfois, la cause des erreurs est extérieure au service chargé du dépouillement. Banister (1980, p. 2) fait observer que, « lorsque nous savons qu'une forte proportion des personnes d'un sous-groupe particulier n'ont pas répondu à une question donnée du recensement, cela signifie qu'elles n'ont pas bien compris la question, ou qu'elles n'ont pas eu envie de coopérer aux opérations de recensement ». Par conséquent, les taux de non-réponse relatifs aux sous-groupes devraient figurer selon elle dans les données enregistrées au titre du recensement et dans les tableaux publiés. Les bureaux nationaux de recensement/statistique ont à présent davantage de possibilités pour conserver ces données sur disque compact ou sur d'autres supports pour les besoins des chercheurs.

55. Il est de plus en plus évident que l'importance plus ou moins grande des opérations de vérification automatique ne saurait se substituer au soin particulier apporté à la collecte des données du recensement. Les bureaux nationaux de recensement/statistique n'ignorent pas qu'à partir d'un certain stade les vérifications automatiques sont non seulement limitées, mais deviennent inefficaces : le processus de vérification ajoute plus d'erreurs qu'il n'en corrige. Le fait de modifier une rubrique de recensement et celui d'y apporter des corrections ne sont nullement équivalents. Par conséquent, l'équipe de vérification doit se concerter pour définir la portée et les limites précises de son action.



56. Les vérifications et les imputations peuvent certes améliorer ou non la qualité des données, mais un jeu de données dont les erreurs ont été éliminées facilite considérablement son utilisation et l'analyse. Le processus commence par la conception même du questionnaire du recensement. Les démographes et les autres spécialistes du sujet déterminent généralement sa teneur, souvent en consultation avec des groupes d'utilisateurs, mais en définitive, les données des recensements ne sont pas élaborées « en priorité à destination de démographes puristes, mais pour un public beaucoup plus vaste d'universitaires, de responsables et de non-spécialistes » (Banister, 1980, p. 17). Il est néanmoins essentiel de réaliser un recensement dont les données invalides et incohérentes ont été éliminées, lorsque la crédibilité du bureau de recensement/statistique est en jeu. Comme le signale Banister, « les services de recensement peuvent citer des cas d'articles humoristiques parus dans la presse ou de lettres indignées adressées aux responsables du recensement par des citoyens, au sujet de tableaux statistiques faisant état de grands-pères âgés de 3 ans, et d'usagers de trains de banlieue parfaitement fictifs ».

57. Le problème est de savoir jusqu'où il faut aller pour obtenir un jeu de données de bonne qualité. Tel qu'indiqué plus haut, l'apparition des ordinateurs, d'abord des gros ordinateurs, puis des micro-ordinateurs, a permis une automatisation pratiquement complète des opérations de vérification. Dans nombre de bureaux nationaux de recensement/statistique, les démographes sont en fait devenus de fervents partisans des vérifications. Aussi y réalise-t-on à présent de nombreux contrôles de cohérence autrefois difficiles à mettre en place, en particulier ceux impliquant des contrôles inter-enregistrements et interménages. Malheureusement, cette possibilité offerte par les micro-ordinateurs a également entraîné nombre de problèmes, au premier rang desquels figure la survérification.

## 1. Comment la survérification est-elle préjudiciable

58. La survérification a un effet préjudiciable sur le processus de contrôle des données, et ce à plusieurs égards, notamment en termes de respect des délais, de coût et de distorsion des valeurs vraies. De plus, il en résulte un faux sentiment de sécurité quant à la qualité des données. Ces différentes préoccupations sont passées en revue ci-dessous.

### a) Respect des délais

59. Le dépouillement est d'autant plus long que les opérations de vérification effectuées par un bureau national de recensement/statistique sont nombreuses. L'essentiel est de déterminer dans quelle mesure la durée supplémentaire augmente la valeur des résultats du recensement. Les équipes de vérification doivent donc évaluer, tant en permanence qu'à posteriori les avantages nets d'un allongement de la durée de dépouillement et d'une augmentation des moyens affectés au recensement dans son ensemble. Souvent, les gains sont si limités au regard du temps investi qu'il vaut mieux tolérer de petites « anomalies » dans les données que priver les principaux utilisateurs de l'obtention en temps opportun des informations escomptées.

### b) Charges financières

60. De manière analogue, les coûts liés au recensement augmentent avec la durée du dépouillement. Il incombe à chaque bureau national de recensement/statistique de décider, lorsqu'il entreprend des vérifications plus importantes et plus complexes, si

l'augmentation des coûts est justifiée par le surcroît de travail et peut être effectivement prise en charge.

c) **Distorsion des valeurs vraies**

61. Bien que le processus de vérification ait pour objectif d'améliorer la qualité des données, la multiplication et la complexité croissante des corrections risquent d'avoir un effet préjudiciable. Parfois, les équipes de vérification modifient par erreur des données pour différentes raisons : défaut de communication entre le démographe et les spécialistes de traitement des données; erreurs d'un programme sophistiqué particulièrement complexe; ou manipulation à plusieurs reprises d'une donnée dans le cadre d'un contrôle. Les bureaux nationaux de recensement/statistique cherchent à éviter ce type de problème dans la mesure du possible. Comme le montrent Granquist et Kovar (1997), le fait d'imputer l'âge d'un mari et d'une épouse en supposant entre eux une différence d'âge prédéfinie peut s'avérer extrêmement utile, mais risque d'introduire une distorsion artificielle des données lorsque ces cas sont très nombreux.

d) **Un sentiment erroné de sécurité**

62. La survérification des données confère au personnel des bureaux nationaux de recensement/statistique, ainsi qu'aux autres utilisateurs, un faux sentiment de sécurité, surtout lorsque les services en question s'abstiennent d'adopter et de documenter des mesures d'assurance qualité. En outre, les tableaux obtenus feront apparaître certaines anomalies indépendamment de l'importance du travail réalisé par l'équipe de vérification; aussi importe-t-il d'avertir les utilisateurs de la présence éventuelle de petites erreurs. Tel est particulièrement le cas maintenant que de nombreux pays diffusent des échantillons de microdonnées. Les bureaux nationaux de recensement/statistique ne souhaitant pas diffuser des données susceptibles de nuire au processus de planification, il faut veiller tout particulièrement à ce que toutes les variables décisives fassent l'objet de vérifications adéquates et soient utilisables à des fins de planification. Par exemple aucun bureau de recensement/statistique ne souhaiterait diffuser des microdonnées ou des tableaux comportant des données inconnues de sexe ou d'âge. En revanche, des variables telles que les caractéristiques relatives à l'incapacité ou à l'aptitude à lire ou à écrire sont parfaitement utilisables avec des vérifications moins importantes. Bien que les tableaux croisés puissent faire apparaître certains défauts de cohérence parce que les bureaux nationaux de recensement/statistique ne peuvent vérifier tous les couples de variables, les équipes de vérification devraient contrôler les principales combinaisons. Lorsque des défauts de cohérence sont mis en évidence, des procédures de correction doivent avoir été mises au point.

## **2. Traitement des données inconnues**

63. L'équipe de vérification doit décider à un stade précoce de la planification du recensement comment traiter les données « non spécifiées » ou inconnues. Tel qu'indiqué plus haut, les colonnes ou les lignes de données inconnues dans les tableaux étant dépourvues aussi bien de valeur informative que d'utilité, dans la plupart des pays les planificateurs du recensement préfèrent imputer les données en question. En outre, certaines valeurs informatives non valides ne peuvent être traitées comme « blancs » mais doivent être résolues. Par exemple, « l'Amérique du Sud » en tant que lieu de naissance doit être remplacée par un pays (tel que le Pérou). L'équipe de vérification doit donc décider du traitement systématique des données inconnues.

### 3. Modifications parasites

64. Bien que cela soit possible, d'ordinaire les bureaux nationaux de recensement/statistique n'utilisent pas de modèles lorsqu'ils définissent leurs règles de vérification. Les équipes de vérification devraient définir des règles adaptées aux caractéristiques réelles de la population ou de l'habitat. Toutes les données recueillies doivent satisfaire à ces règles. Par exemple, il est possible de spécifier par une série de règles que l'enfant d'un chef de ménage doit avoir au moins 15 ans de moins que lui. Toutefois, il peut s'agir d'un enfant « par alliance » et non biologique. Cet enfant peut être l'enfant biologique du conjoint, mais non du chef de ménage, auquel cas la différence d'âge peut être inférieure à 15 ans. Puisque les planificateurs dans la plupart des pays ne réservent pas un traitement distinct pour les enfants et les enfants par alliance, si dans les circonstances ci-dessus les règles de vérification ont pour effet de corriger l'âge de l'enfant, il risque d'apparaître des défauts de cohérence en matière de niveau de formation, de participation à la population active et dans d'autres domaines. Aussi convient-il de tester ce principe afin d'examiner les résultats avant de l'appliquer pleinement.

### 4. Détermination de tolérances

65. L'équipe de vérification doit définir des niveaux de tolérance pour chaque donnée et parfois pour des combinaisons de données. Les niveaux de tolérance indiquent le nombre des réponses non valides et incohérentes admis avant que les équipes de vérification adoptent des mesures correctives. Ainsi, pour la plupart des rubriques d'un recensement, un petit pourcentage des personnes interrogées ne fourniront pas de réponses « acceptables », pour une raison ou une autre. Pour certains types de données, par exemple, l'âge et le sexe, généralement combinées à d'autres types de données à des fins de planification, le niveau de tolérance peut être très bas. Lorsque le pourcentage de réponses manquantes ou incohérentes est faible (inférieur à 1 ou 2 %), aucune règle judicieuse de vérification ne risque d'affecter l'utilisation des données. En revanche lorsque le pourcentage est élevé (5 à 10 % ou davantage selon la situation), une imputation simple ou même complexe risque d'entraîner une distorsion dans les résultats du recensement.

66. Pour réduire au minimum la proportion des réponses manquantes, les bureaux nationaux de recensement/statistique devraient veiller à ce que les recenseurs fassent tout leur possible pour obtenir les données sur le terrain. Si un pays donné décide qu'il n'a pas besoin d'une précision très élevée pour certains types de données, telles que la capacité à lire ou à écrire, ou les caractéristiques d'invalidité, le niveau de tolérance relatif à des données peut être nettement plus élevé. Parfois, les équipes de vérification peuvent corriger des données comportant des erreurs trop nombreuses en renvoyant les agents recenseurs sur le terrain, en effectuant de nouvelles entrevues téléphoniques ou en mettant à profit leur connaissance d'un secteur de recensement particulier. Souvent, toutefois, il s'avère trop coûteux de retourner sur le terrain ou de réaliser d'autres opérations de suivi, et le bureau national de recensement/statistique peut convenir soit de ne pas utiliser les données en question, soit de les utiliser seulement en leur associant des notes de mise en garde.

67. La question se pose de savoir qui doit déterminer le niveau de tolérance pour une information. L'équipe de vérification, notamment les démographes et les informaticiens, est parfois tenue de prendre une décision à cet égard. Les spécialistes doivent utiliser ces informations de façon durable et le fait de veiller à obtenir les données de la meilleure qualité possible est un enjeu professionnel. Quant aux spécialistes

du traitement des données, ils peuvent estimer ne pas être effectivement en mesure de mettre au point des programmes de vérification appropriés afin de ramener le niveau de tolérance à des valeurs acceptables ou encore que les données proprement dites n'autorisent pas l'utilisation efficace d'un programme quelconque avec un niveau de tolérance donné.

## 5. Enseignements du processus de vérification

68. Une fois terminée la vérification des données, il faut garder une trace des analyses détaillées des effets positifs et négatifs obtenus, afin d'améliorer la qualité tant de l'enquête et du recensement actuel que des enquêtes et des recensements futurs. L'équipe de vérification doit s'employer constamment à déterminer ce qui fonctionne correctement et ce qui ne fonctionne pas. Elle doit en outre établir si les aspects du processus qui fonctionnent correctement peuvent être améliorés et simplifiés, pour que les utilisateurs disposent plus facilement des résultats. Les bureaux nationaux de recensement/statistique seront d'autant plus en mesure de corriger les erreurs qu'ils les auront détectées plus tôt au cours du recensement.

## 6. Assurance qualité

69. L'assurance qualité joue un rôle important dans tous les recensements. Par conséquent, des mécanismes formels d'assurance qualité doivent incontestablement être mis en place pour surveiller le bon déroulement de la phase de vérification et d'imputation automatiques. Les analyses rétrospectives, les mesures de performance et les statistiques de diagnostic sont des outils essentiels pour l'analyse de la qualité des vérifications, comme pour la rapidité du dépouillement (Granquist et Kovar 1997; Statistique Canada, 1998).

## 7. Coût des opérations de vérification

70. Le présent manuel peut aider les pays à réduire les coûts élevés en termes de temps et de moyens matériels à engager pour mener à bien la vérification et l'imputation des données des recensements ou des enquêtes. Comme le font observer Granquist et Kovar (1997, p. 418), même « dans les années 90, les opérations de vérification s'avèrent pratiquement aussi coûteuses qu'elles l'étaient dans les années 70, bien qu'elles aient été dans une large mesure rationalisées grâce à la mise à profit permanente de l'évolution technologique ». Pour la plupart des pays, les opérations de vérification absorbent un temps et des moyens financiers disproportionnés : aussi chacun d'eux doit-il déterminer quelle doit être la rentabilité de son investissement. D'après les mêmes auteurs, le coût de vérification des enquêtes sur les ménages représentait quelque 20 % du budget total des recensements réalisés dans le monde au début des années 90.

71. Les vérifications excessives risquent de retarder la publication des résultats. Bien que les agents nationaux chargés des recensements ou des enquêtes aient peut-être uniquement des confirmations anecdotiques de ce type d'expérience en matière de recensement, il ressort d'une étude de Pullum, Harpham et Ozsever (1986) que la vérification automatisée de l'enquête mondiale sur la fécondité (*World Fertility Survey*) a contribué à retarder la publication des résultats d'environ un an. Les bureaux nationaux de recensement/statistique auraient peut-être intérêt à consacrer leurs moyens financiers en priorité à l'amélioration de la qualité de la phase de dénombrement du recensement ou de l'enquête.

## 8. Imputation

72. L'imputation désigne le processus consistant à résoudre les problèmes concernant les réponses manquantes, non valides ou incohérentes détectées par les vérifications. L'imputation a pour effet de modifier une ou plusieurs des réponses ou informations manquantes d'un ou plusieurs enregistrements en cours de vérification, afin d'assurer l'obtention de résultats plausibles, dotés d'une cohérence interne. Toutefois, il est généralement impossible de résoudre tous les problèmes dans la phase initiale en raison de préoccupations touchant au fardeau de réponse, au coût et au respect des délais. Les opérations d'imputation traitent ensuite des autres anomalies décelées par la vérification, puisqu'il est souhaitable d'établir un fichier complet et cohérent contenant des données imputées. Le meilleur travail d'imputation est réalisé par les membres de l'équipe qui ont pleinement accès aux microdonnées et qui détiennent de bonnes informations complémentaires :

- a) Après imputation, le fichier corrigé ne doit pas différer considérablement du fichier vérifié où la présence d'erreurs a été constatée. Aussi, l'imputation d'un nombre minimal de variables est généralement la meilleure méthode, puisqu'elle protège autant que possible les données fournies par les personnes interrogées. Cela suppose implicitement (ce qui n'est pas toujours vrai dans la pratique) qu'une personne interrogée est plus susceptible de faire seulement une ou deux erreurs qu'un nombre plus important;
- b) Après imputation, le fichier obtenu doit satisfaire aux exigences de toutes les vérifications;
- c) Les équipes de vérification devraient identifier comme telles les valeurs imputées, tandis que les méthodes et les sources d'imputation devraient être clairement identifiées;
- d) Afin de pouvoir évaluer l'importance des opérations d'imputation et leurs répercussions, l'équipe de vérification devrait conserver les valeurs imputées et non imputées des champs correspondants des enregistrements des données.

## 9. Archivage

73. Une partie des opérations d'assurance qualité relatives aux recensements ou aux enquêtes consistent à documenter toutes les tâches, puis à archiver les informations ainsi réunies. Les bureaux nationaux de recensement/statistique doivent conserver les fichiers de données vérifiées et non vérifiées à des fins d'analyse ultérieure. Certaines procédures, telles que de nombreuses formes de numérisation, conservent automatiquement l'image saisie initialement. De manière analogue, immédiatement après la saisie des lots de données, il convient de concaténer les données ainsi obtenues et de les conserver en vue d'analyses ultérieures éventuelles. Toutefois, quelle que soit la procédure employée, il importe d'archiver les exemplaires originaux des fichiers non vérifiés. En fait, des copies des données non vérifiées devraient être gardées dans plusieurs endroits au sein du Bureau de statistique, ainsi que dans d'autres endroits du pays et à l'extérieur du pays.

74. La documentation établie doit être suffisamment complète pour permettre aux organisateurs des recensements ou des enquêtes de reconstituer les mêmes opérations à une date ultérieure afin de garantir la compatibilité avec l'enquête ou le recensement considéré. Les opérations effectuées et les résultats obtenus doivent nécessairement être reproductibles. Enfin, les données non vérifiées comme les données vérifiées

doivent être enregistrées en plusieurs endroits, et des mesures appropriées doivent être prises pour garantir leur disponibilité permanente au fil des ans.

75. Comme il a été indiqué ailleurs, une partie de la documentation traite des deux types de rapports de vérification. Le premier rapport fournit des statistiques sommaires en indiquant les nombres et les pourcentages d'erreurs (sur la base de dénominateurs appropriés, tels que le nombre total d'unités d'habitation, la population totale, la population en âge de travailler, le nombre de femmes adultes, etc.). Le second rapport contient au moins un exemple de la structure par « cas », avec l'enregistrement non vérifié des ménages ou des habitations, la liste des erreurs et leur résolution pour l'unité d'habitation ou les personnes vivant dans l'unité en question, et les données vérifiées sur l'unité d'habitation ou le ménage.

76. Les deux ensembles d'erreurs devraient être fournis aux niveaux géographiques logiques, certainement pour les grandes subdivisions administratives; par ailleurs, le fait de fournir des listes d'erreurs aux niveaux géographiques inférieurs pourrait aider à cibler les problèmes liés à la formation des recenseurs, au contrôle de qualité et aux autres questions liées au dénombrement.

## Chapitre II

# Mise en œuvre des opérations de vérification

77. Le présent chapitre donne une vue d'ensemble globale des applications du processus de vérification et d'imputation. Il offre un cadre pour le déroulement général de la vérification d'un recensement ou d'une enquête à partir de données brutes ou de données saisies, en vérifiant la structure et la teneur de ces données, présentant un ensemble de données vérifiées<sup>5</sup>. Certains exemples ont été choisis pour illustrer la nature des problèmes posés par des données non vérifiées et les raisons pour lesquelles il est plus intéressant de disposer de données dûment vérifiées. Ce chapitre examine les problèmes de saisie et de codage liés aux opérations préliminaires de vérification. Il traite également des principaux aspects des opérations de vérification automatisée et des principes généraux en matière, par exemple, de contrôle de validité et de cohérence. Deux types généraux de vérification automatisée, à savoir les techniques d'imputation statique (*cold deck*) et d'imputation dynamique (*hot deck*), font l'objet d'un examen détaillé.

78. Qu'un jeu de données d'un recensement fasse l'objet d'une lecture optique ou d'une saisie manuelle, il suit un certain processus général. L'équipe de vérification des données du recensement dispose au départ de données non vérifiées. Le plus souvent, toutes les données ont été précodées par l'agent recenseur ou par le personnel de bureau, de sorte que l'ensemble de données est prêt pour la vérification structurelle. Parfois, il faut une opération pour convertir les données saisies par lecture optique en une autre forme lisible sur machine pour la vérification, selon le programme de vérification utilisé. Dans d'autres cas, les données ainsi saisies ont également besoin d'une deuxième opération de codage automatique afin qu'il soit possible de répondre à certaines rubriques telles que le lieu de naissance, l'activité économique et la profession.

79. Dans un cas comme dans l'autre, les données non vérifiées devraient apparaître sous une forme qui permette aux programmes informatiques d'effectuer les vérifications structurelles (décrites de façon plus détaillée au chapitre III). La vérification structurelle nécessite un contrôle afin de s'assurer que toutes les grandes subdivisions administratives sont présentées en ordre géographique ou numérique, et que, au sein de chaque grande subdivision administrative, chaque petite subdivision administrative est présentée en ordre géographique ou numérique. Ensuite, que chaque localité est représentée, en ordre géographique ou numérique, dans chaque petite subdivision administrative. Cette procédure se poursuit jusqu'au niveau géographique le plus bas. Comme on le verra au prochain chapitre, il importe d'établir des procédures appropriées pour faire en sorte que chaque unité d'habitation apparaisse et n'apparaisse qu'une fois dans le jeu de données.

80. La vérification structurelle doit également garantir que tous les types d'enregistrements soient présents en temps voulu et qu'aucun type d'enregistrement ne soit répété s'il ne doit pas l'être. Ainsi, pour un recensement de la population et de l'habita-

5 Lorsque ce manuel a été rédigé à l'origine pour les recensements de 2000, presque tous les pays entraient leurs données manuellement. Aujourd'hui, la plupart des pays les entrent par lecture optique, parfois avec suivi par saisie manuelle. Le présent manuel s'efforce de tenir compte de la lecture optique pour les vérifications de structure et de teneur. Alors même qu'il était en préparation, de nouvelles technologies apparaissaient, rendant nécessaire l'utilisation conjointe d'assistants pour les données personnelles et de l'Internet pour la collecte des données et une vérification interactive [voir, par exemple, Ireback (2000) pour l'utilisation d'Internet en Suède]. De même que les pays en voie de développement technologique se heurtaient à des difficultés pour la lecture optique au début des années 2000, nombre de pays estiment aujourd'hui que l'utilisation d'assistants pour les données personnelles a encore besoin d'améliorations.

tion, il convient de décider qui des enregistrements de la population ou de l'habitation doivent venir en premier, et de respecter ensuite ce qui a été convenu pour tout l'ensemble de données. Le plus souvent, seuls les enregistrements de l'habitation doivent être présentés, de sorte qu'il faut savoir que faire des enregistrements excédentaires; les programmeurs doivent également fournir des enregistrements de l'habitation aux ménages qui n'en ont pas. De même, des enregistrements de la population doivent être fournis pour les unités d'habitation occupées (généralement définies comme telles dans l'enregistrement de l'habitation) mais pas pour les unités inoccupées.

81. En ce qui concerne la structure en place, il est important de noter qu'inévitablement, la vérification structurelle sera réexaminée lors de la vérification de la teneur des données, et souvent au delà, car des anomalies peuvent apparaître aux différentes phases du processus de recensement. C'est là une procédure normale, à laquelle il faut s'attendre; par conséquent, il faut prévoir du temps, du personnel et des équipements lors de la conception d'ensemble du système.

82. La vérification de la teneur des données peut alors commencer. Chaque rubrique d'un recensement de la population et de l'habitation doit être examinée de façon individuelle et, généralement aussi, conjointement avec d'autres rubriques, afin de déterminer la validité attribuée à chacune et la meilleure combinaison des différentes rubriques. Les chapitres IV et V traitent des différentes données sur la population et sur l'habitation mentionnées dans la publication de l'Organisation des Nations Unies intitulée *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision*.

83. Lorsque la teneur des données a été vérifiée, un jeu complet de données vérifiées devrait être établi. Les données non vérifiées devraient être conservées en plusieurs endroits sûrs, et les éléments importants non vérifiés (voire tous les éléments non vérifiés) devraient figurer à la fin des divers types d'enregistrements. Une fois encore, il importe de souligner que, lors de l'établissement des tableaux, il pourrait s'avérer nécessaire de réexaminer également les vérifications de la teneur, afin que tout problème particulier résultant de certaines tabulations croisées puisse être résolu.

84. La vérification des recensements et des enquêtes a pour objet de mettre en évidence les omissions et les défauts de cohérence des enregistrements de données. Les opérations d'imputation visent à introduire les corrections nécessaires. La vérification définit les procédures spécifiques adaptées au traitement des omissions et des données inacceptables. Les opérations d'imputation modifient les données non valides et résolvent les incohérences détectées dans l'ensemble des informations introduites. Il en résulte un fichier de microdonnées vérifiées en vue des divers types d'exploitation, contenant toutes les données élémentaires pertinentes relatives à chaque unité de logement et à chaque personne recensée.

85. Il importe néanmoins de signaler à nouveau que la vérification, si poussée soit-elle, ne saurait remplacer un bon travail de dénombrement. Le processus de vérification est efficace lorsque les imputations s'appliquent aux omissions et aux incohérences aléatoires. Toutefois, si des erreurs systématiques affectent la collecte de données, la vérification ne peut améliorer la qualité des données quelle que soit la complexité des procédures mises en œuvre. Par ailleurs, le choix des sujets étudiés est d'une importance décisive pour la qualité des données recueillies. Les personnes interrogées doivent être désireuses et capables de fournir les informations adéquates. Il est donc parfois nécessaire d'éviter des sujets susceptibles d'éveiller des craintes, des préjugés locaux ou des superstitions, ou encore de poser des questions trop complexes et trop difficiles pour permettre à la personne interrogée « moyenne » de répondre facilement dans le contexte d'un recensement de la population. La formulation précise de chaque



question nécessaire à l'obtention de la réponse la plus fiable possible sera obligatoirement fonction du contexte national particulier et devrait être soigneusement testée au préalable. Pour obtenir des données de meilleure qualité, il importe donc au plus haut point que les bureaux nationaux de recensement/statistique engagent des ressources suffisantes consacrées au recensement.

86. Afin de mettre en place la phase de vérification automatisée du processus, l'équipe de vérification établit à cet effet des instructions ou des spécifications rédigées par écrit, des tables de décision, des organigrammes et des pseudocodes. Le pseudocode est un ensemble d'instructions écrites ou de spécifications tel qu'indiqué à la figure 8.

87. Les organigrammes permettent aux spécialistes du sujet de mieux comprendre les différentes relations entre les variables et facilitent l'élaboration des instructions de vérification. L'annexe IV présente des exemples d'organigrammes. Les spécialistes du sujet rédigent les instructions de vérification en collaboration avec les informaticiens, en décrivant l'opération à effectuer pour chaque élément d'information; ces instructions doivent être claires, concises et dépourvues d'ambiguïté puisqu'elles constituent le point de départ du progiciel de vérification.

88. Toute l'équipe de vérification du recensement, tant les démographes que les informaticiens, doit être parfaitement familiarisée avec le traitement et l'analyse des données démographiques. En effet, un personnel non qualifié risque d'introduire dans le recensement des erreurs et des biais supplémentaires.

## A. Considérations de codage

89. Comme indiqué précédemment, pendant une bonne partie de la seconde moitié du XX<sup>e</sup> siècle, les pays ont introduit manuellement leurs données. Bien qu'aujourd'hui la plupart des pays saisissent les données nécessaires à leurs recensement pas lecture optique, souvent ils continuent de le faire par saisie manuelle. Même lorsque les formulaires sont passés par lecteur optique, certaines variables ont encore besoin d'être transcrites de mots en chiffres. Le processus de production de chiffres et de données alphanumériques lisibles par machine est appelé codage.

90. Certains progiciels de vérification peuvent facilement accepter et exploiter des données alphanumériques, mais la plupart d'entre eux se heurtent à des difficultés de classement, de récapitulation et d'établissement de pourcentages ou de valeurs médianes, etc., lorsqu'ils ont à traiter des données non numériques.

91. Il convient d'éviter dans toute la mesure possible les codes composés entièrement de caractères alphabétiques ou de caractères alphabétiques combinés à des chiffres (c'est-à-dire des données alphanumériques). Lorsque les formulaires sont soumis à lecture optique, les données alphanumériques ne posent pas de problèmes importants; en revanche, pour beaucoup de progiciels, leur utilisation exige une forte dose de manipulation ou, du moins, réclame une attention considérable. De nombreux programmes de vérification ne peuvent traiter les caractères alphabétiques que s'ils sont placés entre guillemets ou isolés de quelque autre manière.

92. Lorsqu'ils conçoivent un programme de codage, les agents chargés du recensement ou de l'enquête doivent tenir compte du rendement de chaque investissement en temps, en énergie et en ressources financières. Les considérations de codage sont relativement insignifiantes pour les petits pays ou les petites enquêtes, car le volume de traitement nécessaire est beaucoup plus modeste que pour un recensement.

En outre, les données qui sont saisies par lecture optique ne souffrent pas sensiblement des problèmes que posent des colonnes d'informations supplémentaires.

93. En revanche, lorsqu'un recensement ou une enquête utilise deux colonnes pour la rubrique *liens*, par exemple, au lieu d'une, la lecture optique introduit des erreurs qui seraient absentes s'il n'y avait qu'une colonne d'informations; autrement dit, avec les codes 1 à 9, le lecteur optique peut capter un caractère alphabétique, ou un blanc, ou un signe parasite converti en caractère lisible. Toutefois, ces problèmes sont automatiquement résolus lors de la vérification, comme il est indiqué plus avant dans cette publication.

94. En revanche, quand il y a deux colonnes, par exemple, codes 1 à 10, on court le risque d'introduire tout un nouvel ensemble d'erreurs. Au lieu des valeurs légales 1 à 9, les valeurs entrées pourraient aller de 0 à 99, en plus des caractères alphabétiques, des blancs ou des signes parasites mentionnés précédemment. Lorsque les vérificateurs reçoivent une valeur de 13, il leur faut commencer à prendre des décisions stratégiques sur ce qu'ils doivent faire de cette valeur. Devait-il s'agir d'un 3, et le 1 est-il donc erroné? Ou la valeur devait-elle être 10, ce qui voudrait dire que le 3 est donc inexact? Dans la plupart des cas, les spécialistes du sujet fournissent des spécifications de vérification pour cette rubrique, mais la présence de ces valeurs augmente automatiquement la durée et la complexité de la vérification et pourrait diminuer la valeur du jeu de données final.

95. L'un des problèmes les plus courants, examiné spécifiquement plus avant dans cette publication, se pose au sujet des rubriques de la série fécondité. Nombre de pays recueillent à présent des informations sur les enfants du ménage, les enfants se trouvant ailleurs et les enfants décédés, et parfois, sur le nombre total de ces enfants et sur le sexe de chacun. Ainsi, les pays pourraient avoir jusqu'à 12 éléments d'information. La question qui se pose ici est de savoir combien de chiffres devraient être utilisés pour chacune de ces rubriques. Lorsque l'on utilise deux colonnes, la valeur pour les garçons pourrait se situer n'importe où entre 0 et 99; lorsque l'on n'utilise qu'une colonne, ces chiffres ne peuvent varier que de 0 à 9. Toutefois, comme il est extrêmement peu probable qu'une femme ait plus de 9 garçons dans sa famille, le fait d'utiliser deux chiffres introduit une forte probabilité de collecte de signes parasites ou d'erreurs de lecture, par exemple le risque qu'au lieu de lire un chiffre compris entre 0 et 9, on n'aboutisse à un total de 91 enfants au lieu de 01.

96. Par conséquent, pour les garçons et les filles présents au sein du ménage, se trouvant actuellement ailleurs ou décédés, une seule colonne serait probablement ce qui conviendrait le mieux. En revanche, pour le total des enfants présents au sein du ménage, se trouvant actuellement ailleurs ou décédés, il pourrait être préférable d'utiliser deux colonnes. Cela dépend pour beaucoup des niveaux de fécondité enregistrés dans le pays. À l'occasion, une famille pourra effectivement avoir plus de neuf personnes dans une catégorie particulière, mais, comme toujours dans un recensement, il appartiendra au bureau de statistique de trouver le juste équilibre entre erreurs et données exactes.

97. Pour les variables ordinales, on peut considérer la série suivante de codes de liens :

- 1 Chef de ménage (ou chef de famille)
- 2 Conjoint
- 3 Enfant
- 4 Frère ou sœur
- 5 Parent

- 6 Petit-fils ou petite-fille
- 7 Autre membre de la famille
- 8 Personne n'appartenant pas à la famille

Dans la plupart des pays, cet ensemble de codes standard couvre la majorité des liens. Certains pays ajoutent un code « 0 » pour le chef de famille et peuvent donc ajouter une dixième catégorie aux autres.

98. Ces codes peuvent être utilisés pour déterminer la composition du ménage, comme il est indiqué à l'annexe I sur les variables calculées. Toutefois, nombre de pays, en particulier ceux qui sont frappés par l'épidémie de VIH/sida, ont besoin d'informations beaucoup plus détaillées que ne permettent d'en obtenir ces codes. Ces pays peuvent avoir besoin d'informations spécifiques sur les enfants par alliance, les beaux-parents, les grands-parents, les nièces et les neveux, et ainsi de suite. Dans ce cas, le bureau de statistique a besoin de codes additionnels à deux chiffres pour accomplir sa mission.

99. Quand un pays décide d'opter pour de multiples colonnes, il lui faut également décider de la façon de les utiliser. Dans l'exemple présenté ci-dessus, on suppose que les codes de lien seront séquentiels. Toutefois, lorsque la décision est prise d'utiliser deux colonnes, les spécialistes du sujet peuvent décider d'attribuer une signification particulière aux colonnes elles-mêmes. Par exemple :

- 10 Chef de ménage
- 11 Conjoint
- 12 Frère et sœur
- 13 Conjoint du frère ou de la sœur
- 21 Enfant
- 22 Enfant adoptif
- 23 Enfant d'un autre lit
- 24 Nièce/neveu
- 31 Parent
- 32 Beau-parent
- 33 Oncle/tante
- 41 Petit-enfant
- 77 Autre membre de la famille
- 88 Personne n'appartenant pas à la famille
- 90 Population d'institutions

100. Cette formule affecte des codes dans la première colonne selon la génération : 1 pour la génération du chef de ménage; 2 pour la génération suivante; 3 pour la génération précédente; 4 pour la deuxième génération après celle du chef de ménage, etc., puis le deuxième chiffre selon le type de lien de parenté. Bien que ces valeurs puissent être utiles dans la reconstitution de la famille, le personnel du bureau de statistique et certains utilisateurs extérieurs peuvent les trouver incommodes.

101. Néanmoins, ce type de codage devrait être envisagé pour certaines variables sociales et économiques. Pour ce qui est de l'appartenance ethnique, par exemple, le premier chiffre indiquerait le principal groupement tribal ou ethnique ou un groupement tribal ou ethnique secondaire. Ben entendu, lorsqu'il y a plus de 10 groupements secondaires, il conviendrait d'utiliser deux chiffres dans la première colonne.

102. De même, pour les rubriques nécessitant trois ou quatre chiffres, par exemple pour la profession ou l'activité économique, le premier chiffre indiquerait la

catégorie générale de profession ou d'activité économique principale; le deuxième chiffre, la catégorie secondaire de la profession ou de l'activité économique; et le troisième chiffre, la profession ou l'activité économique précise. Comme la plupart des systèmes de codage internationaux comportent déjà des niveaux inscrits dans les codes, le bureau de statistique n'a pas à effectuer de travail supplémentaire.

103. Lorsque les bureaux nationaux de recensement/statistique élaborent des listes de codes pour les programmes de vérification et pour l'exploitation ultérieure des données, ils pourraient désirer établir des codes communs pour certaines rubriques. Par exemple, dans nombre de pays, les codes d'endroits (lieu de naissance, lieu de naissance des parents, lieu de résidence précédent, lieu de travail), la langue, la race et l'appartenance ethnique et la nationalité sont très semblables. Un ensemble commun de codes pour « le lieu » pourrait être conçu sous la forme d'un code à trois chiffres où le premier chiffre représenterait le continent, le deuxième, la région et le troisième, le pays en question. Les bureaux nationaux de recensement/statistique peuvent également utiliser des codes numériques de pays conçus par des organisations internationales telles que la Division de statistique de l'Organisation des Nations Unies (Nations Unies, 1999). Un ensemble commun de codes applicables à des variables étroitement liées peut aider à réduire les erreurs de codage et faciliter le travail des informaticiens lors de la vérification. De tels codes permettent également aux informaticiens, le cas échéant, d'utiliser une donnée introduite sous une rubrique pour en déterminer une autre.

104. La structure du codage peut faciliter le processus de codage ainsi que les opérations ultérieures lors de la vérification, de l'exploitation et de l'analyse. Pour les grands pays comptant de nombreux immigrants, des codes fondés sur le continent, la région et le pays, et des codes ou des chiffres différents attribués à chacun, seraient préférables au simple établissement d'une liste.

105. La figure 1 offre des exemples de codes communs pour certaines rubriques telles que le lieu de naissance, la nationalité, la langue et l'appartenance ethnique. Pour les Philippines, les codes pour les personnes parlant l'Ilokano et le Tagalog diffèrent du code général appliqué aux langues du pays. Selon la situation particulière du pays, ces codes peuvent également différer les uns des autres. Bien que la langue anglaise ait un code unique, elle est parlée par plus d'un groupe ethnique. C'est pourquoi les codes utilisés au Canada pour le lieu de naissance, la nationalité et l'appartenance ethnique diffèrent légèrement de ceux utilisés aux États-Unis. Pour les personnes nées en France, ayant la nationalité française, parlant le français et appartenant au groupe ethnique français, on utilise le même code. Ainsi, si l'une de ces informations manque et si l'équipe de vérification décide que la solution est appropriée, l'informaticien peut prendre le code de l'une des autres données.

106. Si diverses questions figurant sur un questionnaire ne sont pas indépendantes les unes des autres, le personnel national de recensement/d'enquête ne devrait probablement pas les poser toutes. L'équipe de vérification devrait décider, dans chaque cas, quand utiliser d'autres questions directement et quand utiliser d'autres variables disponibles.

Figure 1

#### Exemples de codes communs pour certaines rubriques

Groupe	Lieu de naissance	Nationalité	Langue	Ethnicité
France/français	10	10	10	10
Espagne/espagnol	20	20	20	20

Groupe	Lieu de naissance	Nationalité	Langue	Ethnicité
Amérique latine	25	25	20	25
Philippines/philippin	30	30	30	
Ilokano			32	
Tagalog			32	
Angleterre/anglais	40	40	40	40
Canada	50	50	40	50
États-Unis	52	52	40	52

107. Un autre problème se pose quand les définitions diffèrent d'un recensement à l'autre (ou entre un recensement et une enquête) pour certaines variables telles que le travail ou l'appartenance ethnique. Le bureau national de recensement/statistique doit décider de la façon de tenir compte de ces changements pour indiquer les tendances. Si les données initiales non vérifiées sont disponibles, les informaticiens peuvent introduire les changements dans les vérifications appropriées et reprendre toutes les vérifications.

108. Par exemple, un pays européen peut utiliser un code unique pour le pays d'origine pour tous les pays d'Asie du Sud lorsque seuls quelques cas sont identifiés. Toutefois, en raison de l'évolution des mouvements migratoires, l'enquête ou le recensement peut exiger des codes distincts pour l'Inde, le Bangladesh, le Pakistan, Sri Lanka et les autres pays d'Asie du Sud jusqu'au stade du traitement.

## B. Vérification manuelle et vérification automatisée

109. La vérification manuelle d'un recensement peut prendre des mois ou des années, et donner lieu ainsi à de nombreuses possibilités d'erreur humaine. De fait, il s'agit d'une alternative qui laisse à désirer par rapport à la vérification automatisée, notamment en raison de l'impossibilité de créer ou de reconstituer une analyse rétrospective du processus de correction manuelle. Les vérifications informatisées ou automatisées réduisent les délais et limitent les risques d'erreur humaine. Les vérifications tant informatisées que manuelles ont pour effet de vérifier la validité d'une valeur introduite par comparaison à une valeur acceptable, tandis que les programmes informatisés contrôlent également la cohérence de la valeur introduite par rapport aux valeurs connexes. Enfin, et avant tout, les vérifications automatisées permettent d'établir une analyse rétrospective et sont donc reproductibles, contrairement aux vérifications manuelles.

110. Dans les premières années de la saisie par ordinateur, aucune vérification des valeurs introduites n'était possible, ce qui veut dire que toutes les corrections devaient être effectuées à la main, dans le cadre des opérations de codage et de vérification entreprises par le bureau, ou devaient s'inscrire dans le cadre des opérations informatiques qui suivaient la saisie manuelle des données. Les nouveaux progiciels comportent des fonctions de vérification qui permettent d'assurer que les données non valides ne peuvent être introduites que sciemment par l'opérateur et que les incohérences peuvent être signalées pour être corrigées par l'opérateur ou par un programmeur. Lorsque la lecture optique a gagné du terrain, cette évolution s'est répétée : durant les premières années de la lecture optique, aucune vérification n'était possible lors de la saisie des données; toutefois, récemment, des vérifications de validité et des conversions et recodages de données ont été incorporés aux systèmes de lecture optique.

111. Lorsque les recensements et les enquêtes recueillent de grandes quantités de données, il n'est pas toujours possible de se référer aux documents originaux pour corriger les erreurs. Même si l'on dispose des questionnaires d'origine, les données consignées sur ces derniers risquent parfois d'être erronées ou incohérentes. Or, un système informatisé de vérification et d'imputation permet de corriger ou de modifier immédiatement des données erronées et d'établir des rapports pour toutes les erreurs décelées et pour toutes les modifications effectuées. Il convient de planifier soigneusement les vérifications informatisées afin de préserver le temps dont dispose le personnel pour les autres activités de traitement des données. Le passage d'importantes quantités de données dans un système informatique peut être long, mais il prend moins de temps que les corrections manuelles.

112. Les corrections manuelles sont réalisées sous différentes formes. Considérons l'exemple simple d'une erreur concernant le sexe d'une personne recensée : un contrôleur vérifie le travail d'un agent recenseur et constate une erreur évidente, telle que l'enregistrement de la mention « homme » dans le cas d'une personne prénommée « Marie ». S'il corrige le sexe et inscrit la mention « femme », le contrôleur effectue une correction manuelle. Si le contrôleur ne corrige pas le questionnaire mais l'envoie au contraire au bureau de recensement local, le problème en question peut y être constaté et corrigé manuellement. Au bureau central, le défaut de concordance entre le prénom et le sexe peut être décelé au cours du codage et corrigé manuellement par le codeur; sinon, le personnel chargé d'introduire les données relatives au questionnaire peut observer le défaut de concordance entre le prénom et le sexe et apporter les corrections nécessaires avant la saisie.

113. Toutefois, si l'erreur n'est pas décelée et si l'opérateur de saisie introduit le code « homme », plusieurs procédures peuvent alors être exécutées à ce stade. Pour les éléments d'information liés au sexe, comme ceux recueillis par le bloc de questions sur la fécondité, le programme de vérification peut indiquer de fait qu'il s'agit d'un homme auquel est associée une information sur la fécondité et produire un message à cet effet au moment de la saisie des données. L'opérateur pourrait ensuite examiner le questionnaire et établir qu'il s'agit effectivement d'une femme, et effectuer manuellement la correction. Sinon, au cas où le bureau national de recensement/statistique utilise un programme de vérification indépendant de la saisie, le programme informatique pourrait indiquer cette personne comme étant un homme auquel sont associées des données sur la fécondité. Les employés de bureau peuvent retrouver le questionnaire original dans les boîtes, l'examiner et établir que la personne interrogée, prénommée Marie, a été enregistrée par erreur en tant qu'« homme ». À ce stade, le personnel de bureau peut remonter cette information à l'opérateur, lequel peut reprendre l'enregistrement et le corriger manuellement.

114. Cet exemple fait apparaître conjointement les avantages et les inconvénients de la vérification manuelle. À chacune des étapes décrites ci-dessus, un employé du recensement a pu constater l'erreur, le défaut de concordance entre le prénom et le sexe, et faire la correction nécessaire. Toutefois, il existe dans les bureaux nationaux de recensement/statistique qui utilisent le système de vérification manuel des agents qui vérifient à chaque étape cette concordance. Il faut consacrer à cette tâche une somme de travail considérable, pour des résultats vraisemblablement peu différents (en particulier les résultats d'ensemble), de ceux obtenus avec la consigne de n'effectuer aucune vérification manuelle.

115. Au départ, la seule façon de corriger un ensemble de données consistait à introduire manuellement les modifications. Certains pays jugent encore peu pratique d'utiliser un système de correction automatique, et font appel au système manuel, à l'une des étapes évoquées ci-dessus. Si le jeu de données étudiées est limité ou si le fac-

teur temps n'est pas décisif, ou encore si l'on dispose d'une main d'œuvre importante, alors le choix des corrections manuelles sera viable dans de nombreux cas. Il a ainsi pour avantage, si l'information recueillie sur le questionnaire est complète et exacte, et si l'examen de ce dernier permet vraiment de résoudre les incohérences, d'entraîner probablement une amélioration limitée de la qualité du recensement ou de l'enquête (l'équipe de vérification est alors tenue de supposer par exemple que le prénom « Marie » n'est pas « Gary »; qu'en présence de données sur la fécondité, l'information effectivement recueillie correspond à cette personne; et que cette information n'a pas été recueillie de façon erronée). En fait, les procédures de vérification et d'imputation améliorent rarement la qualité de la collecte des données. Elles modifient seulement certains éléments.

116. Parfois, l'examen d'un questionnaire en vue d'une correction manuelle s'avère infructueux, l'information recherchée étant absente, pour une raison ou une autre. Parfois, une personne ne souhaite pas indiquer son âge et laisse en blanc la case correspondante, auquel cas, l'étude du questionnaire ne résoudra pas la question. Ensuite, l'équipe de vérification doit prendre une décision quant à la façon de résoudre le problème. Pour procéder à une correction manuelle, le bureau national de recensement/statistique doit soit attribuer la valeur « inconnu », soit utiliser un ensemble de valeurs susceptibles d'être attribuées à la rubrique « âge ».

117. La correction manuelle altère inévitablement la qualité et la cohérence des réponses si la personne interrogée n'est pas contactée. Elle prend plus de temps et coûte davantage. Les ordinateurs ne sont pas exposés à la fatigue et sont donc plus rapides; en outre, ils ne souffrent pas de problèmes personnels qui peuvent s'avérer préjudiciables au maintien de la qualité ou de la cohérence des données; et, dans la plupart des cas, ils permettent de réduire le coût du traitement. Aussi, la plupart des pays utilisent-ils à présent une forme ou une autre de correction automatique.

118. Les omissions et les incohérences observées dans les réponses réduisent la qualité des données et rendent problématique la présentation sur tableaux de résultats du recensement aisément compréhensibles. Certains utilisateurs préfèrent mettre en tableaux les réponses manquantes et incohérentes regroupées sous la rubrique « non indiqué », tandis que d'autres préfèrent répartir ces cas proportionnellement aux données cohérentes indiquées. D'autres encore préconisent des règles d'imputation de réponses probables aux réponses manquantes ou incohérentes. Grâce aux ordinateurs, l'imputation de réponses à partir des autres informations figurant dans le questionnaire, ou sur la base d'informations indiquées pour une personne ou une unité de logement présentant des caractéristiques similaires, est une opération parfaitement réalisable et ne présentant pas de difficulté.

119. Puisque l'ordinateur offre la possibilité d'examiner de nombreuses caractéristiques, le processus de vérification doit mettre à profit cette capacité. Les procédures de vérification impliquant un grand nombre de caractéristiques connexes peuvent donc conduire à imputer davantage de réponses vraisemblables que ne le permettrait une vérification simple. En revanche, une opération de vérification mal conçue risque d'aboutir à la production de données de recensement de qualité médiocre. L'équipe de vérification devrait être constituée de spécialistes expérimentés des différentes disciplines concernées, ainsi que d'informaticiens. Les membres de cette équipe devraient procéder soigneusement au choix des variables à examiner dans les tests de cohérence afin de définir les spécifications de vérification et d'imputation. De plus, parmi les résultats du programme devrait figurer le pourcentage de réponses modifiées ou imputées : les analystes seront ensuite en meilleure position pour juger de la qualité des données; par exemple, un fort pourcentage d'imputations constituerait un avertissement, invitant à une utilisation prudente des résultats.

120. L'analyse rétrospective des vérifications fait apparaître les modifications apportées à chaque variable. Elle permet de tracer l'historique des réponses depuis la réception des données et tout au long du processus de vérification et d'imputation.

## C. Principes de correction des données

121. Manuel ou automatique, le processus de vérification devrait s'employer à rendre les données aussi représentatives que possible de la situation réelle en éliminant les omissions et les réponses non valides et en modifiant les réponses incohérentes.

### Encadré 2

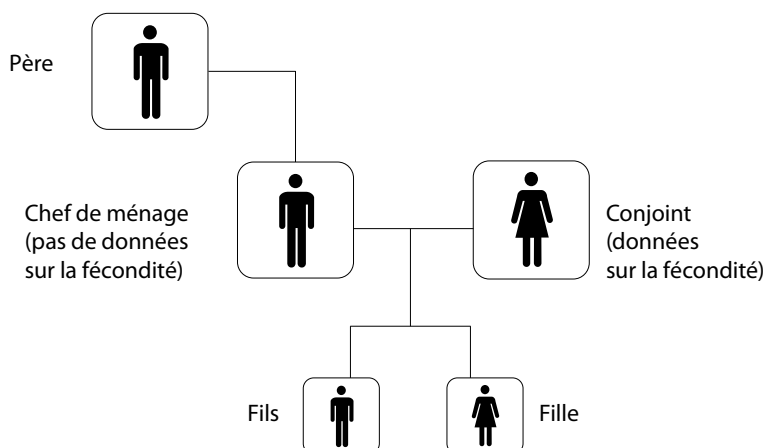
#### Principes importants de correction des données

Lors de la mise au point des procédures de vérification, il peut être utile de ne pas perdre de vue les suggestions suivantes en matière de correction des données :

- ☞ Effectuer le plus petit nombre possible de modifications des données initialement enregistrées;
- ☞ Éliminer les incohérences évidentes des réponses;
- ☞ Établir des valeurs correspondant aux éléments inexacts ou manquants sur la base d'autres éléments d'information concernant le logement, la personne ou les autres personnes du ménage ou d'un groupe comparable, conformément aux procédures spécifiées. Dans certains cas, il convient d'utiliser pour certains éléments d'information la catégorie de réponse « non indiquée ».

122. Considérons le diagramme suivant (figure 2) relatif à un ménage particulier. Il décrit un ménage et fait apparaître des relations et des données de sexe cohérentes. Le chef du ménage est un homme et ne fait pas l'objet de données sur la fécondité. Le conjoint est du sexe féminin et les informations appropriées de fécondité lui sont associées.

Figure 2  
Exemple de ménage fictif, avec indication du lien avec le chef de ménage, du sexe et de la fécondité de ses membres





123. Dans bien des cas, toutefois, les informations sont incohérentes. Il convient alors de répondre aux questions suivantes : quel devrait être le processus de vérification dans le cas d'un ménage faisant l'objet de réponses incohérentes ? Comment doit procéder l'équipe de vérification si le chef de ménage et le conjoint sont signalés comme étant de sexe masculin, comme sur la figure 3 ? Autrefois, le principe de vérification généralement adopté aurait consisté à supposer que la première personne dans un couple est un homme, en particulier s'il s'agit du chef de ménage, et que la deuxième personne, c'est-à-dire le conjoint, est une femme.

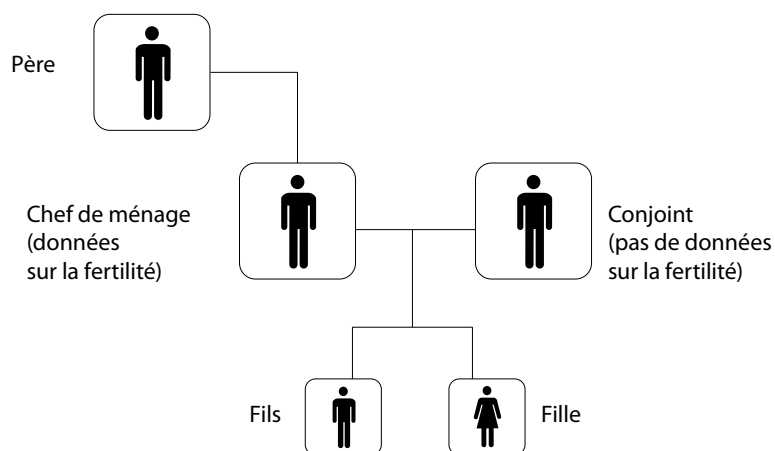
124. Si en l'occurrence le chef de ménage s'avérait être la femme et non le mari, alors la règle de vérification adoptée serait erronée et le bureau national de recensement/statistique enregistrerait en définitive quatre erreurs :

- a) Erreur concernant le sexe du chef de ménage;
- b) Erreur concernant le sexe du conjoint;
- c) Perte de données sur la fécondité relatives au chef de ménage;
- d) Attribution erronée de données sur la fécondité au conjoint masculin.

Il s'agit évidemment d'une procédure de vérification incorrecte.

125. En revanche, après avoir constaté l'identité du chef de ménage et du conjoint, une procédure de vérification correcte contrôle ensuite les données sur la fécondité relatives à ces deux personnes. Puisque seul le chef de ménage comporte des données sur la fécondité, il s'agit de la personne interrogée de sexe féminin. Les règles de vérification concernant ces rubriques sont alors correctement appliquées.

Figure 3  
Exemple de ménage : chef de ménage et conjoint du même sexe



126. L'exemple de la figure 4 illustre également ce problème. Pour la plupart des pays, l'âge de procréation est compris entre 15 et 49 ans. Considérons le cas d'une femme déclarant avoir un enfant à l'âge de 52 ans, d'après les données directes déduites du numéro de ligne indiqué pour la mère de l'enfant, ou d'après la différence d'âge calculée (la différence d'âge entre la mère et l'enfant ne devrait probablement pas dépasser 50 ans, quoique pour les enfants adoptifs, elle puisse être plus grande). L'équipe de vérification doit décider si la différence d'âge est acceptable ou si elle doit être modifiée, par le changement de l'un des deux âges en question. Si l'on augmente la différence d'âge acceptable et si d'autres femmes déclarent avoir des enfants à des

âges plus avancés, des anomalies plus nombreuses risquent d'être introduites dans le jeu de données en cas d'indications erronées concernant l'âge proprement dit. Là encore, l'équipe de vérification doit décider du caractère approprié des âges déclarés pour différentes variables.

Figure 4  
Exemple de ménage avec indication des âges de certains membres

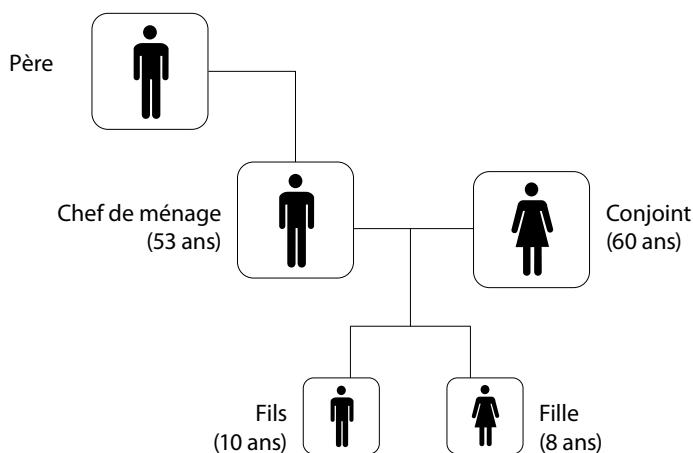
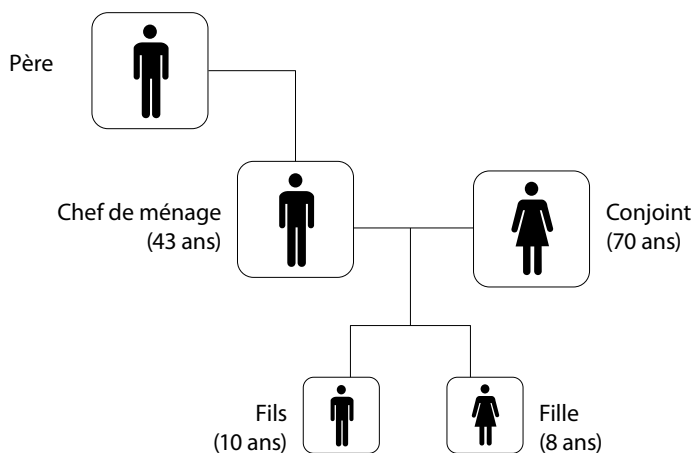


Figure 5.  
Exemple de ménage présentant des incohérences potentielles des âges déclarés



127. La figure 5 illustre un autre scénario possible. Supposons que les vérifications détectent le cas d'une femme de 70 ans avec des enfants de 8 et 10 ans, tel qu'indiqué à la figure 5. Cette situation est possible, le mari pouvant avoir eu les enfants avec une précédente épouse. Dans ces conditions, les enfants sont rattachés au chef de ménage et non à l'épouse en tant que telle, bien qu'il soit sans doute plus probable que les opérateurs aient fait une erreur en tapant le chiffre 7, croyant taper le chiffre 4 correspondant à 40 ans. Pour une raison ou une autre, supposons que les spécialistes du sujet demandent aux informaticiens de modifier l'âge de la mère ou de l'enfant lorsque

leur différence d'âge dépasse 50 ans. Or, une demande de ce type entraîne des vérifications encore plus compliquées. Puisque l'âge de la femme est de 70 ans et celui du premier enfant de 10 ans, l'équipe de vérification doit choisir la personne dont l'âge doit être modifié; elle pourrait décider de porter à 20 ans l'âge du premier enfant, ce qui résoudrait le problème dans ce cas particulier, ou encore, de modifier l'âge du conjoint. Il reste néanmoins à résoudre le problème posé par l'âge du deuxième enfant, qui doit aussi faire l'objet d'une vérification.

128. Si l'on considère uniquement les âges de la mère et d'un enfant, une imputation attribuerait l'âge de façon aléatoire et donnerait des résultats exacts dans environ 50 % des cas. Toutefois, si l'on considère également l'âge du mari, l'équipe de vérification aurait davantage tendance à modifier l'âge de l'épouse, compte tenu de cette information complémentaire. Cette modification aurait pour effet de renforcer la compatibilité des âges des différents membres de la famille.

## D. Contrôle de validité et de cohérence

129. Parmi les principales exigences des opérations de vérification figure l'élimination des valeurs non valides. En outre, il faut assurer la cohérence des réponses fournies pour chacune des rubriques connexes à l'intérieur de chaque enregistrement et d'un enregistrement à l'autre. Les données non valides sont celles qui sont inacceptables pour des raisons techniques ou des raisons d'esthétique. Par exemple, des codes ne sont admis pour le sexe que pour les hommes et les femmes. Toute autre valeur serait inacceptable et devrait être remplacée par la mention « inconnu » ou par une valeur applicable à l'un des deux sexes; comme la plupart des pays entreprennent leurs travaux de planification et de formulation de politique sur la base du sexe pour de nombreuses variables, la mention de valeurs inconnues dans le jeu de données compliquerait les efforts en vue d'obtenir les valeurs uniques nécessaires à ces travaux. De même, des dépouillements présentant des incohérences telles que « murs de chaume et toiture en béton », « femmes de 13 ans ayant 20 enfants », « titulaires d'un doctorat âgés de trois ans » donneraient au bureau de statistique une apparence d'incompétence, même si les rares incohérences restaient sans incidence sur les efforts de planification du pays.

130. Les opérations d'imputation devraient prendre en compte toutes les informations disponibles concernant les variables connexes au même moment, dans toute la mesure possible, et pas nécessairement de façon séquentielle, par rapport aux variables connexes. Dans certains cas, cependant, la vérification peut procéder à un contrôle de cohérence avant de déterminer la validité d'une donnée. Lorsque l'imputation attribue une valeur sur la base d'un contrôle de cohérence, elle doit comparer la valeur à la donnée d'origine pour déterminer s'il y a eu effectivement modification; sinon, la donnée initiale reste inchangée.

### 1. Méthode de vérification descendante

131. Cette procédure commence par la première rubrique à vérifier (le « haut »), généralement la première variable du questionnaire, puis passe successivement aux différentes rubriques, jusqu'à la dernière. La méthode habituelle consiste à tenir compte en premier des taux de réponse et de l'importance relative des diverses rubriques. Les vérifications commencent généralement par le sexe et l'âge, en raison de leur importance, en particulier pour l'imputation dynamique. Bien que la méthode descendante

ne tienne pas pleinement compte des relations entre les éléments d'information, il s'agit toutefois d'une procédure adéquate pour mener à bien la vérification.

132. Au cours de la vérification, certaines opérations modifient la valeur d'une rubrique à plusieurs reprises. Cette procédure risque d'introduire une ou plusieurs erreurs dans le jeu de données. Une valeur imputée peut en effet ne pas être cohérente par rapport aux autres données. Même lorsque les variables sont traitées de façon séquentielle, une variable particulière doit être vérifiée en tenant compte, si possible, simultanément de toutes les autres. Par exemple, s'il est imputé en fonction de l'âge de la mère, l'âge d'un enfant risque d'être incompatible avec le nombre d'années de scolarité déclaré ou d'années passées dans la circonscription. Dans ce cas, l'âge fera l'objet d'une nouvelle imputation jusqu'à ce qu'il soit compatible avec les autres données. Un âge imputé constitue une variable intermédiaire jusqu'à ce qu'une valeur finale soit attribuée. Lors de la création de données vérifiées, les variables intermédiaires imputées ne doivent pas être enregistrées en tant que modifications tant qu'elles n'ont pas été définitivement attribuées.

133. Bien que le programme de vérification puisse accepter la présence d'un « blanc » ou une valeur dite « non indiquée », pour certaines rubriques et dans certaines conditions, les informations connexes peuvent fournir des valeurs pour la plupart des rubriques laissées en blanc ou auxquelles des valeurs erronées ont été attribuées. Les valeurs obtenues de cette façon peuvent être individuellement correctes ou non. Cependant, la puissance et la vitesse de l'ordinateur pour comparer différentes valeurs enregistrées permettent de déterminer les valeurs de remplacement qui rendent compte de la situation de façon satisfaisante. Les tableaux de valeurs obtenus alors, dans la plupart des cas, présenteront une cohérence parfois supérieure par rapport aux tableaux établis à partir des enregistrements non vérifiés ou des enregistrements dans lesquels le processus d'imputation transforme toutes les réponses inacceptables en leur attribuant le code « non indiqué ».

134. Le programme de vérification doit également effectuer des contrôles structurels (voir chapitre III). Les contrôles doivent porter sur les données de population (chapitre IV) comme sur les données d'habitation (chapitre V). En outre, les procédures de vérification devraient vraisemblablement créer une ou plusieurs variables recodées dans l'enregistrement particulier nécessaire à la mise en tableaux, tel qu'indiqué à l'annexe I.

135. Il est extrêmement important d'éviter les vérifications circulaires, qui consistent à modifier une ou plusieurs rubriques puis, à un stade ultérieur, à les rétablir à leur valeur initiale. Le présent manuel fait observer par la suite qu'il incombe au personnel d'exécuter à plusieurs reprises le programme de vérification pour être sûr d'avoir entièrement contrôlé toutes les rubriques. Il est possible de définir des règles de vérification qui modifient les données lors d'un premier passage mais qui, lorsqu'elles sont appliquées aux données modifiées lors d'un second passage, rétablissent la configuration initiale. Cette procédure risque de se répéter à l'occasion de traitements multiples. L'équipe de vérification doit éviter d'introduire des règles de ce type dans le processus opératoire.

## **2. Méthode de vérification portant sur plusieurs variables**

136. La méthode « descendante » de vérification des recensements et des enquêtes, présentée dans la section 1 ci-dessus, ne donne pas toujours les meilleurs résultats (qui se rapprochent le plus de la distribution réelle des variables). Tel qu'indiqué,

la méthode descendante, appliquée sans prendre les précautions qui s'imposent, pose souvent des problèmes de vérification.

137. Une autre méthode à envisager est celle des vérifications portant sur plusieurs variables, qui est fondée sur le système de Fellegi-Holt. Cette méthode exige certes davantage de compétences en matière de traitement de données et une puissance de calcul plus importante, mais elle permet d'obtenir des résultats sans doute plus proches de la « réalité ». L'annexe V intitulée « Méthodes d'imputation » présente différents types de vérifications portant sur plusieurs variables. Ce système comporte nécessairement l'établissement d'une série de propositions « vraies » et la définition du lien entre les variables. Ensuite, le programme applique chaque proposition aux données concernant le ménage afin de vérifier si elles sont toutes exactes. Pour toute proposition erronée, le programme conservera la trace, pour chaque rubrique, des valeurs non valides ou des incohérences. Une fois tous les tests effectués, le système de vérification et d'imputation doit déterminer la meilleure façon de modifier l'enregistrement, pour qu'il subisse avec succès toutes les vérifications. Les équipes de vérification optent généralement pour une stratégie de modification minimale et modifient le plus petit nombre possible de variables afin d'obtenir un enregistrement acceptable.

138. Les onze propositions de la figure 6 donnent un exemple de règles susceptibles d'être appliquées dans le cadre d'une vérification à plusieurs variables de certaines caractéristiques démographiques. Dans cet exemple, le chef de ménage doit avoir 15 ans au moins. Pour des programmes de vérification généraux, il serait plus indiqué de faire figurer un âge minimal de  $x$  années,  $x$  étant la valeur minimale déterminée pour le pays considéré. Les propositions définies dans cet exemple sont axées sur différentes variables primaires importantes, telles que le lien avec le chef de ménage, le sexe, l'âge, l'état matrimonial et la fécondité. Puisqu'il s'agit de variables étroitement liées, les équipes de vérification auraient intérêt à les examiner conjointement, afin de définir la méthode la plus efficace de contrôle des données. Il convient de signaler en l'occurrence que toutes les variables sont certes importantes, mais que certaines jouent un rôle plus décisif du point de vue de la présentation des données.

139. La figure 6 présente un cas simple où, pour une raison quelconque, le chef de ménage et le conjoint sont du même sexe; en l'occurrence ils sont tous deux du sexe masculin, et l'un d'entre eux est présenté avec une donnée sur la fécondité. Il est très clair que le sexe est erroné (comme l'indique la note au bas du tableau) et que le sexe indiqué pour la personne présentant une donnée sur la fécondité devrait être le sexe féminin.

Figure 6  
Exemple de règles de vérification à variables multiples  
de certaines caractéristiques démographiques

N°	Règle	Lien avec le chef de ménage		Âge	État matrimonial		Fécondité
		Sexe					
1	Le chef de ménage doit avoir 15 ans au moins						
2	Le conjoint doit avoir 15 ans au moins						
3	Un conjoint doit être marié						
4	En présence d'un conjoint, le chef de ménage doit être marié						

**Note :** les « 1 » indiquent que deux données ou plus sont incohérentes. Par exemple, à la rubrique 5, comme le chef de ménage et le conjoint sont du même sexe, la vérification indique une erreur pour le lien avec le chef de ménage et le sexe, et un 1 apparaît dans les cases correspondantes.

N°	Règle	Lien avec le chef de ménage		Âge	État	
		Sexe			matrimonial	Fécondité
5	En présence d'un conjoint, le chef de ménage et le conjoint ne doivent pas être du même sexe	1	1			
6	Les individus de moins de 15 ans ne sont jamais mariés					
7	Aucune donnée sur la fécondité n'est associée aux personnes du sexe masculin				1	1
8	Aucune donnée sur la fécondité n'est associée aux personnes du sexe féminin de moins de 15 ans					
9	Les données sur la fécondité doivent toujours être indiquées dans le cas des personnes du sexes féminin de 15 ans ou plus					
10	Un enfant doit toujours être plus jeune que le chef de ménage					
11	Un parent doit toujours être plus vieux que le chef de ménage					
Totaux		1	2			1

140. Dans l'exemple de la figure 7, les deux conjoints appartiennent à la même population que ceux de la figure 6. Ils sont signalés l'un et l'autre comme étant de sexe masculin. En l'occurrence, la procédure de vérification est simple et évidente. La variable pour laquelle les erreurs les plus nombreuses ont été dénombrées est celle qui sera contrôlée en premier. À la figure 7, le programme de vérification met en place la procédure d'imputation relative à la variable « sexe » puisque, d'après les données de la figure 6, cette variable comporte le plus grand nombre d'erreurs eu égard : 1) au lien avec le chef de ménage et au sexe; et 2) à la fécondité et au sexe. Lorsque le programme de vérification contrôle les données sur la fécondité et constate qu'une information sur la fécondité est associée au chef de ménage et non au conjoint, la procédure d'imputation attribue le sexe féminin au chef de ménage. Enfin, lorsque l'équipe de vérification contrôle à nouveau toutes les erreurs dénombrées, et si toutes les propositions irréfutables sont exactes, aucune vérification supplémentaire n'est nécessaire.

Figure 7

**Exemple comportant un chef de ménage et un conjoint de même sexe dans un jeu de données non vérifiées**

Personne	Lien avec le chef de ménage	Sexe	Nombre d'enfants mis au monde
<b>Données non vérifiées</b>			
1	Chef de ménage	Masculin	03
2	Conjoint	Masculin	BLANC
<b>Données après vérification portant sur la variable sexe</b>			
1	Chef de ménage	Féminin	03
2	Conjoint	Masculin	BLANC

141. Les prescriptions techniques concernant cette vérification peuvent être formulées tel qu'indiqué à la figure 8. Si les données sur la fécondité sont précisées dans les deux cas, le programme de vérification fonctionnera. Toutefois, la vérification n'est évidemment pas complète puisqu'elle considère uniquement les cas dans lesquels

les données sur la fécondité sont indiquées et exactes, tant pour le chef de ménage que pour le conjoint.

Figure 8

**Exemple de spécification de vérification visant à corriger les valeurs indiquées pour la variable sexe, en langage symbolique**

---

Si SEXE du CHEF DE MÉNAGE = SEXE du CONJOINT  
 Si FÉCONDITÉ du CHEF DE MÉNAGE n'est pas en blanc  
     Si FÉCONDITÉ du CONJOINT est en blanc  
         (si la variable SEXE du chef de ménage n'est pas déjà mise à la valeur « féminin ») attribuer la valeur SEXE = féminin  
         (et si la variable SEXE du conjoint n'est pas déjà mise à la valeur « masculin ») attribuer la valeur SEXE = masculin  
     Autrement *Une autre correction s'impose... car les deux conjoints ont le même sexe et comportent l'un et l'autre des données sur la fécondité!*  
         [La correction à apporter pourrait être d'utiliser le sexe du précédent chef de ménage, de changer la variable sexe du chef de ménage, ou d'utiliser les ratios attribués aux sexes de tous les chefs de ménage pour produire une réponse appropriée, etc.]  
     Terminé - si  
 Terminé - si  
 Autrement *Tel est le cas où les données sur la fécondité du chef de ménage sont laissées en blanc*  
     Si FÉCONDITÉ du CONJOINT n'est pas en blanc  
         (si la variable SEXE du chef de ménage n'est pas déjà mise à la valeur « masculin ») attribuer la valeur SEXE = masculin  
         (si la variable SEXE du conjoint n'est pas déjà mise à la valeur « féminin ») attribuer la valeur SEXE = féminin  
     Autrement *Une autre correction s'impose parce qu'alors AUCUN DES DEUX conjoints ne comporte de données sur la fécondité!*  
         [La correction à apporter pourrait être d'utiliser le sexe du précédent chef de ménage, de changer la variable sexe du chef de ménage, ou d'utiliser les ratios attribués aux sexes de tous les chefs de ménage pour produire une réponse appropriée, etc.]  
     Terminé - si  
 Terminé - si  
 Terminé - si

---

142. La figure 9 ci-dessous correspond à un exemple de procédure de vérification appliquée à un chef de ménage de sexe féminin âgé de 13 ans, qui est veuve mais a trois enfants, d'après les données saisies. L'application des règles de vérification contenues dans le programme donne les résultats suivants :

Figure 9

**Exemple d'analyses de vérification à plusieurs variables, dans le cas d'une veuve très jeune avec trois enfants**

N°	Règle	Lien avec le chef de ménage	Sexe	Âge	État matrimonial	Fécondité
1	Le chef de ménage doit avoir 15 ans au moins	1		1		
2	Le conjoint doit avoir 15 ans au moins					
3	Un conjoint doit être marié					
4	En présence d'un conjoint, le chef de ménage doit être marié					
5	En présence d'un conjoint, le chef de ménage et le conjoint ne doivent pas être du même sexe					
6	Les individus de moins de 15 ans ne doivent jamais être mariés			1	1	
7	Aucune donnée sur la fécondité n'est associée aux personnes du sexe masculin					

N°	Règle	Lien avec le chef de ménage		État		Fécondité
		Sexe	Âge	matrimonial		
8	Aucune donnée sur la fécondité n'est associée aux personnes du sexe féminin de moins de 15 ans	1	1			1
9	Les données sur la fécondité doivent toujours être indiquées dans le cas des personnes du sexe féminin de 15 ans ou plus					
10	Un enfant doit toujours être plus jeune que le chef de ménage					
11	Un parent doit toujours être plus vieux que le chef de ménage					
Totaux		1	1	3	1	1

143. Considérons à nouveau le cas d'une veuve de 13 ans chef de ménage ayant trois enfants. La première vérification, décrite par la règle n° 1 — chef de ménage âgé de 15 ans au moins — indique une erreur puisque le chef de ménage a moins de 15 ans. Comme elle a 13 ans, le chiffre 1 apparaît dans les cases indiquant le « lien avec le chef de ménage » et « l'âge », car il y a incohérence entre ces deux variables. Elle n'est pas un conjoint, par conséquent, les règles 2 et 3 ne sont pas applicables. Les règles 4 et 5 ne sont pas non plus applicables pour la même raison, elles ne s'appliquent qu'au conjoint. En revanche, comme aux termes de la règle 6, les individus de moins de 15 ans ne doivent jamais être mariés, et l'on est ici en présence d'une veuve âgée de 13 ans, cette règle est violée. Dans la mesure où la règle 7 s'applique aux personnes du sexe masculin, elle n'est pas applicable. En vertu de la règle 8, aucune donnée sur la fécondité ne doit être indiquée pour les personnes du sexe féminin de moins de 15 ans; dans la mesure où l'on est en présence d'une telle personne ayant des enfants (c'est-à-dire, présentant un chiffre de fécondité), cette règle est violée. Les règles 9, 10 et 11 ne s'appliquent pas à la personne en question.

144. D'après la série de propositions vraies, le plus grand nombre d'erreur concerne la variable âge, et celle-ci doit être la première à faire l'objet de modifications. Après modification de la variable âge, on procède à de nouveaux contrôles, et la vérification sera terminée si les corrections introduites résolvent toutes les incohérences. Sinon, le programme procède à la vérification de la variable qui se classe au second rang pour le nombre d'incohérences.

## E. Méthodes de correction et d'imputation des données

145. Tel qu'indiqué plus haut, les données recueillies dans tous les recensements et toutes les enquêtes comportent des blancs, qui correspondent aux mentions « non indiqué » ou « inconnu » ou aux informations manquantes pour une autre raison. La présence de données non valides est due par ailleurs aux erreurs de la personne interrogée, de l'agent de recensement ou de l'opérateur de saisie. Les méthodes utilisées pour introduire les corrections nécessaires dépendent du type de données considéré. Dans la plupart des cas, il est possible d'attribuer aux données en cause des codes valides, tout en étant raisonnablement assuré de leur exactitude en utilisant les réponses fournies sous d'autres rubriques et contenues dans l'enregistrement du ménage ou de la personne ou dans ceux d'autres ménages ou d'autres personnes.

146. Ce manuel présente deux techniques informatisées de correction des données erronées. La première est celle de l'imputation statique (méthode *cold deck*, c'est-



à-dire à donnée fixe) utilisée principalement pour les données manquantes ou inconnues. La seconde est celle de l'imputation dynamique ou méthode *hot deck* (c'est-à-dire à donnée variable) applicable aussi bien aux données manquantes qu'aux données incohérentes ou non valides. Différents progiciels et différents programmes au sein de ces derniers utilisant diverses méthodologies emploient les méthodes *cold deck* et *hot deck* de différentes façons, illustrées dans les annexes.

## 1. Technique d'imputation statique ou *cold deck*

147. Selon la technique d'imputation statique ou *cold deck*, le programme de vérification attribue une réponse déterminée à un élément d'information manquant à partir d'un ensemble de valeurs prédéterminées, ou impute la réponse suivant la règle de proportionnalité à partir d'une distribution de réponses valides. Dans la méthode dite *cold deck*, le programme n'actualise pas le jeu initial de variables. Ainsi, les valeurs utilisées ne diffèrent pas de celles de la matrice statique initiale, une fois terminé le traitement des enregistrements concernant la première, la deuxième, la dixième ou toute autre personne. Les valeurs initiales permettent d'effectuer les imputations nécessaires pour toute donnée manquante.

148. L'imputation statique est une méthode stochastique, ainsi que l'imputation dynamique, mais les valeurs utilisées restent inchangées dans le temps. Cette approche est décrite à l'annexe V.

149. Parfois, l'imputation statique repose sur l'utilisation d'un rapport, et attribue les réponses selon des proportions prédéterminées. Afin d'illustrer ce type de répartition proportionnelle des réponses, considérons un tableau de données valides, c'est-à-dire de données correspondant à des réponses effectivement fournies et non manquantes, concernant la durée de travail hebdomadaire des hommes de 33 ans travaillant dans le secteur de l'agriculture : cette durée était de 50 heures par semaine pour 25 % d'entre eux, de 60 heures pour 40 % et de 70 heures pour 35 %. Dans ce cas, les valeurs manquantes ou non valides de la durée de travail hebdomadaire des hommes de 33 ans travaillant dans le secteur de l'agriculture serait remplacée dans 25 % des cas par 50 heures, dans 40 % par 60 heures et dans 35 % par 70 heures. Toutefois, à moins de disposer de données fiables provenant de recensements, d'enquêtes ou d'une autre source, l'application de cette technique exige au préalable l'exploitation des réponses valides fournies au recensement en cours, solution qui n'est pas nécessairement viable du point de vue économique ou pratique.

## 2. Technique d'imputation dynamique ou *hot deck*

150. Une autre méthode utilisable pour éliminer du jeu de données les éléments d'information réputés « inconnus » est celle de l'imputation dynamique ou technique dite de la donnée variable (ou *hot deck*) qui consiste à attribuer des valeurs à des données non disponibles, inconnues, incorrectes ou incohérentes. Le United States Census Bureau a initialement mis au point la méthode, mais d'autres administrations ont depuis lors ajouté des perfectionnements. La technique de l'imputation dynamique utilise une ou plusieurs variables pour évaluer la réponse probable lorsque le jeu de données comporte une réponse non spécifiée (ou parfois plusieurs valeurs non spécifiées). Cette technique est maintenant de plus en plus répandue pour les vérifications des recensements en raison de sa simplicité et du fait de la reproductibilité et de l'élimination des erreurs des résultats. En outre, grâce à l'élimination des valeurs non spécifiées, il est plus facile de mettre en évidence les tendances entre des enquêtes et

des recensements successifs, puisque l'analyste n'est plus tenu de traiter au cas par cas le problème des valeurs inconnues.

151. Dans le cas de l'imputation dynamique, les données connues concernant des individus présentant des caractéristiques similaires déterminent les informations les plus pertinentes à utiliser lorsqu'un (ou plusieurs) élément d'information concernant un autre individu n'est pas connu. Parmi ces caractéristiques figurent le sexe, l'âge, le lien avec le chef de ménage, la situation économique et le niveau d'instruction. La matrice d'imputation proprement dite est constituée d'un ensemble de valeurs que l'on peut comparer à la donne d'un jeu de cartes. Ces matrices permettent d'enregistrer, puis de fournir les informations utilisées en présence de valeurs inconnues. Cette donne est modifiée en permanence par des opérations de mise à jour et/ou « en battant les cartes » suivant la procédure logique de façon à ce que les imputations évoluent au cours du traitement des données, d'où le terme de *hot deck* (donne variable).

152. Les valeurs mises en mémoire dans la donne variable représentent des informations sur les « voisins les plus proches » présentant des informations similaires. Il convient de noter que le voisin le plus proche est généralement le voisin *précédent* le plus proche parce que, surtout dans la méthode descendante décrite ailleurs, les unités d'habitation et les personnes habitant dans ces unités ne sont prises en considération qu'une fois, puis le programme se poursuit. Par conséquent, dans un village, notamment, lorsque l'on ignore que quelqu'un est orphelin de mère, la donne variable contient des informations sur la personne du même sexe et du même âge rencontrée le plus récemment dont la qualité d'orphelin de mère a été avérée. Cette méthode est particulièrement importante dans les pays caractérisés par de vastes mouvements migratoires ou par une forte prévalence du VIH/sida ou d'autres phénomènes donnant lieu à une activité statistique inhabituelle. De même, les caractéristiques concernant l'habitat ont plus de chances d'être similaires à l'intérieur d'un village ou d'un ensemble de villages qu'entre un village ou un ensemble de villages et d'autres régions du pays.

153. À titre d'exemple simple, supposons qu'une valeur unique soit enregistrée en tant que donne. Par exemple, si le sexe d'une personne est non valide pour une raison ou une autre, une valeur initiale arbitraire (de sexe masculin ou féminin) détermine une valeur initiale. La valeur de départ ainsi obtenue est attribuée à la première personne dont la variable « sexe » n'est pas spécifiée. En revanche, en cas de validité de la valeur enregistrée pour la variable sexe de la première personne, cette même valeur est substituée à la valeur de départ. Si ensuite, la valeur de la variable sexe de la deuxième personne n'est pas connue, alors la matrice d'imputation lui attribue la valeur enregistrée. Dans ce cas, le sexe imputé est donc celui de la première personne. En principe, lorsque le programme de vérification établit le caractère acceptable de la valeur d'une rubrique, il place cette valeur dans la matrice d'imputation. Par contre, lorsqu'il établit le caractère inacceptable d'une valeur, le processus d'imputation lui substitue la valeur valide contenue dans la matrice d'imputation.

154. La méthode d'imputation dynamique (*hot deck*) décrite ci-dessus pose notamment le problème suivant : si deux rubriques différentes ont une valeur inconnue, l'attribution de réponses valides risque de ne pas utiliser le même individu donneur. Chaque valeur peut provenir d'une personne « réelle », mais il peut s'agir de personnes différentes. Une meilleure façon de procéder consisterait à attribuer simultanément une valeur aux deux variables, à partir de la même personne. La programmation de ces matrices complexes risque toutefois de présenter certaines difficultés.

155. Les données présentées à la figure 10 ci-dessous portent sur un ménage composé de 10 individus. Les blancs, marqués d'un « X » ou de deux X (« XX ») indiquent des données manquantes. Souvent les chiffres 9 et 99 sont utilisés pour désigner

des informations manquantes, en l'occurrence 9 pour le sexe et 99 pour l'âge. Parfois, cependant, le chiffre 9 remplace une autre valeur réelle, par exemple, dans un nombre limité de codes pour le lien; par conséquent, ces valeurs sont à utiliser avec parcimonie et, en fait, si l'on peut utiliser un « X » ou « .. », c'est probablement ce qui devrait se faire. Il est à noter que, bien que d'autres variables soient utilisables pour effectuer les imputations, par exemple, le niveau d'instruction et la profession, elles ne sont pas prises en compte dans cet exemple succinct.

Figure 10

**Exemple de ménage utilisé pour fournir les données d'entrée d'une imputation dynamique**

Numéro d'identification	Lien avec le chef de ménage	Sexe	Âge
1	1	1	39
2	2	2	35
3	3	1	13
4	3	X	10
5	4	2	40
6	4	1	XX
7	4	2	13
8	5	X	XX
9	5	1	44
10	5	2	36

NOTE : X ET XX = information manquante.

156. Si la valeur initiale utilisée dans la matrice d'imputation SEXARRAY est « sexe masculin » (code = 1), la matrice d'imputation devient SEX = 1.

157. Après traitement des données de la personne n° 1, la valeur reste égale à 1. Toutefois, après traitement des données concernant la deuxième personne, il lui est substitué la valeur 2, puisque la personne considérée est de sexe féminin. La variable sera alors : SEXARRAY = 2.

158. Dès qu'une valeur valide est saisie en ce qui concerne le sexe d'une personne dont les données sont traitées, le code correspondant au sexe de cette personne remplace la valeur stockée dans la matrice d'imputation. Lorsque les données de la troisième personne sont traitées, l'imputation rétablit la valeur 1 (sexe masculin).

159. Dans le cas de la quatrième personne, la valeur de la variable « sexe » est inconnue : le programme de vérification examine la valeur contenue dans la matrice d'imputation qui correspond en l'occurrence au sexe masculin, et remplace la valeur inconnue par la valeur contenue dans la matrice d'imputation. La personne n° 5 est de sexe féminin, et la valeur correspondante remplace la valeur précédente contenue dans la matrice d'imputation provenant de la personne n° 3 (sexe masculin). Ce processus se poursuit jusqu'à la personne n° 8.

160. Le programme de vérification utilise à nouveau la procédure d'imputation et la personne n° 8 se voit attribuer la valeur « sexe féminin », puisqu'il s'agit de la valeur correspondante de la matrice d'imputation provenant de la personne n° 7. Le programme de vérification a utilisé à deux reprises la matrice d'imputation pour déterminer des valeurs : une première fois pour obtenir une valeur « sexe masculin » et une deuxième fois pour obtenir une valeur « sexe féminin ». Comme les fréquences respectives des sexes sont sensiblement identiques, à long terme le processus d'imputation utilise chaque valeur de cette variable, sensiblement une fois sur deux. Après traitement des données concernant chacune des 10 personnes, la variable prendra la valeur : SEXARRAY = 2.

161. Il existe par ailleurs d'autres modalités plus complexes d'utilisation de cette procédure d'attribution d'une valeur à la variable sexe au moyen d'une matrice d'imputation. Par exemple, la vérification peut utiliser le lien avec le chef de ménage et le sexe pour faciliter la détermination de l'âge d'un individu. Considérons la liste partielle suivante de codes attribués au lien avec le chef de ménage :

- 1 = Chef de ménage
- 2 = Conjoint
- 3 = Enfant
- 4 = Autre parent
- 5 = Personne non apparentée

162. Le traitement peut créer pour la variable âge des valeurs initiales pouvant rendre compte de la situation réelle des liens, pour chaque sexe. Ces valeurs ne seront cependant pas très importantes, puisque le programme de vérification les remplacera très certainement avant de les utiliser. De plus, le programme comporte nécessairement l'imputation de nombreuses valeurs, de telle sorte que peu de valeurs initiales ont une incidence sur les tableaux obtenus en définitive. Ces valeurs pourraient se présenter tel qu'indiqué à la figure 11.

Figure 11  
Matrice initiale d'imputation statique de l'âge  
d'après le sexe et le lien avec le chef de ménage

	Lien avec le chef de ménage				
	Chef de ménage	Conjoint	Fils/Fille	Autre parent	Personne non apparentée
	(1)	(2)	(3)	(4)	(5)
Homme (1)	35	35	12	40	40
Femme (2)	32	32	12	37	37

163. Examinons à nouveau le cas des dix personnes du ménage de la figure 10. Puisque la première personne de notre échantillon figure dans la liste en tant que chef de ménage (code = 1) et qu'il s'agit d'un individu de sexe masculin (code = 1), son âge (39) se substitue au premier élément d'information (coordonnées 1,1) au cours de la procédure d'imputation. La « donne » contient alors les valeurs indiquées à la figure 12.

Figure 12  
Exemple de matrice d'imputation dynamique à la suite d'une modification

	Lien avec le chef de ménage				
	Chef de ménage	Conjoint	Fils/fille	Autre parent	Personne non apparentée
	(1)	(2)	(3)	(4)	(5)
Homme (1)	39	35	12	40	40
Femme (2)	32	32	12	37	37

164. La deuxième personne de l'échantillon est le conjoint (code = 2), qui est de sexe féminin (code = 2); aussi, son âge (35) remplace-t-il la valeur figurant à la deuxième ligne et à la deuxième colonne, se substituant aux valeurs de la donne. Les

âges des autres personnes du ménage sont également substitués aux valeurs contenues dans la matrice d'imputation, jusqu'à la personne n° 5.

165. Comme on peut le constater, la précédente procédure d'imputation de la variable sexe a attribué la valeur 1 à la personne n° 4. Puisque le programme de vérification exige l'imputation d'une valeur à la variable sexe, la matrice d'imputation n'est pas actualisée au moyen de l'âge de cette personne. Le programme de vérification procédera à une mise à jour uniquement avec les valeurs tirées des enregistrements dans lesquels le sexe et le lien avec le chef de ménage ont des valeurs initialement correctes dans les deux cas. Pour le traitement des données de la personne n° 6, le programme constate toutefois que la valeur de l'âge n'est pas spécifiée. Il s'agit d'une personne de sexe masculin, figurant parmi les « autres parents » du chef de ménage. Le programme de vérification utilise par conséquent l'élément de la matrice d'imputation relatif aux individus de sexe masculin de la catégorie « autre parent » définie quant au lien avec le chef de ménage (quatrième colonne, première ligne) et attribue la valeur de l'âge pour cette catégorie (« autre parent de sexe masculin » dans ce cas, 40).

166. Aucune donnée de sexe ni d'âge n'est mentionnée dans le cas de la personne n° 8. Le programme de vérification impute à la variable sexe la valeur sexe féminin, puis attribue l'âge en fonction du sexe et du lien avec le chef de ménage (5). Dans ce cas, l'âge attribué est 37 ans.

167. Bien que le programme de vérification ait imputé la valeur de l'âge à partir du lien connu avec le chef de ménage, il a utilisé pour les autres variables une valeur préalablement attribuée au sexe. En l'occurrence, l'usage des valeurs déjà attribuées pour de nouvelles imputations est un exemple de procédure de vérification sujette à caution (voir section 3, *d* ci-dessous). Il serait en effet préférable de chercher d'autres données connues, par exemple l'état matrimonial, pour procéder aux imputations.

168. Après la personne n° 10, les valeurs de la matrice d'imputation sont indiquées à la figure 13. Dans cet exemple, les deux imputations ont utilisé la matrice statique initiale. Les opérations d'imputation n'utiliseront qu'un petit nombre de valeurs initiales, sinon aucune. Dans la majorité des cas, les valeurs attribuées utilisées proviendront de la population recensée.

Figure 13

**Exemple de matrice d'imputation dynamique à la suite de plusieurs modifications**

	Lien avec le chef de ménage				Personne non apparentée
	Chef de ménage	Conjoint	Fils/fille	Autre parent	
	(1)	(2)	(3)	(4)	(5)
Homme (1)	39	35	13	40	44
Femme (2)	32	35	12	13	36

**3. Problèmes posés par l'imputation dynamique (méthode *hot deck*)**

*a)* Considérations géographiques

169. Si le programme de vérification attribue les valeurs manquantes selon la méthode de l'imputation dynamique, il doit alors chercher à utiliser des données classées en fonction de la zone géographique de moindre étendue. Cette procédure doit en principe augmenter la probabilité d'obtention d'une réponse exacte, en raison de l'homogénéité relative normalement observée parmi les gens qui vivent dans la même

unité géographique restreinte, eu égard à leurs différentes caractéristiques, notamment démographiques. Sans homogénéité de la population, il n'y aura pas de corrélation et l'équipe de vérification devra examiner les variables au cas par cas. En outre, comme il est mentionné plus avant, certaines variables ne s'appliquent pas à certaines régions — tel est le cas du chauffage central dans les régions chaudes — et le programme de vérification devrait en tenir compte.

#### **b) Utilisation de rubriques connexes**

170. Avant de recourir à la méthode d'imputation dynamique pour obtenir les valeurs manquantes, il convient de chercher à utiliser les rubriques connexes pour attribuer une valeur susceptible d'être correcte. Par exemple, en cas d'omission concernant l'état matrimonial d'une personne, le programme de vérification établira si la personne en question est associée à un conjoint au sein du ménage. Si tel est le cas, le programme attribuera le code relatif aux personnes mariées sans utiliser la matrice d'imputation. Toutefois, faute de données utilisables de ce type, le programme devra sans doute recourir à une matrice d'imputation.

#### **c) Effet de l'ordre des variables sur les matrices**

171. Les bureaux nationaux de recensement/statistique qui utilisent les matrices d'imputation devraient réfléchir au choix des variables dont ils ont besoin lorsqu'ils établissent l'ordre dans lequel s'effectuent les vérifications. Pour les variables démographiques, les bureaux de recensement souhaiteront procéder d'emblée à la vérification des données de sexe et d'âge, pour pouvoir utiliser ces valeurs dans les autres matrices d'imputation. La vérification générale ne devrait pas utiliser de variables non vérifiées dans les matrices d'imputation, bien que la plupart des progiciels acceptent des colonnes ou des lignes contenant des « valeurs inconnues ». Les taux de réponse et la répartition des attributs au sein des variables faciliteront la détermination des meilleures variables et les attributs les plus utiles au sein de ces variables pour l'établissement des données variables. Les matrices d'imputation ultérieures peuvent utiliser les données dûment vérifiées. Toutefois, dans la mesure du possible, les bureaux de statistique devraient envisager d'exclure les données vérifiées de la matrice d'imputation.

172. Par exemple, si le programme de vérification impute l'âge en fonction du sexe et du lien avec le chef de ménage, les éléments de la rangée correspondante de cette matrice d'imputation (sexe et lien avec le chef de ménage) ne devraient pas être mis à jour en cas d'imputation de l'une ou l'autre des variables sexe et lien avec le chef de ménage. En règle générale, le progiciel de vérification devrait introduire une valeur de l'âge correspondant aux valeurs appropriées des variables sexe et lien avec le chef de ménage uniquement lorsque les données d'âge, de sexe et de lien avec le chef de ménage sont toutes valides et cohérentes. Cependant, l'usage de données corrigées est parfois inévitable pour d'autres raisons. Il importe de souligner que la plupart des pays ne font pas cas de cette suggestion et procèdent à une imputation à partir de valeurs imputées antérieurement. La solution pourrait être d'utiliser des signes d'imputation afin de décourager l'utilisation de données imputées pour aligner un donneur sur une unité erronée.

#### **d) Complexité des matrices d'imputation**

173. Le bureau national de recensement/statistique augmente la probabilité d'obtention d'une valeur « correcte » et cohérente au moyen de la matrice d'imputation en augmentant le niveau de détail de cette dernière. Ainsi, le programme pourrait

imputer l'état matrimonial à partir du seul lien avec le chef de ménage. Toutefois, la probabilité de veuvage ou de divorce augmentant avec l'âge, l'imputation de l'état matrimonial sur la base de l'âge et du lien avec le chef de ménage est donc parfaitement fondée. Partant de l'âge et du lien de la personne considérée avec le chef de ménage, le programme de vérification utilise « l'état matrimonial » contenu dans la matrice d'imputation et provenant d'une personne dotée des mêmes caractéristiques, figurant dans l'enregistrement valide qui précède immédiatement.

174. Or, la procédure décrite ci-dessus risque de créer de nouveaux problèmes. Le bureau national de recensement/statistique vérifie généralement les rubriques du questionnaire dans un ordre fixe, l'âge à la suite de l'état matrimonial, selon une approche descendante. Si tel est le cas, lorsque l'état matrimonial et l'âge ont tous deux été omis, il est impossible de prendre la valeur de la variable « état matrimonial » provenant du fichier qui précède immédiatement et comportant des valeurs identiques de l'âge et du lien avec le chef de ménage<sup>6</sup>. De ce fait, le programme risque de ne pas pouvoir déterminer la catégorie d'âge relative à cet enregistrement. Une autre solution consisterait à prévoir dans la matrice d'imputation une rangée ou une colonne correspondant aux rubriques « non indiqué ». Or, deux arguments vont à l'encontre de ce choix. D'une part, les cases « non indiqué » correspondant à la même combinaison sont si peu nombreuses qu'il serait difficile de mettre à jour la matrice pour la rubrique manquante. D'autre part, il est pratiquement impossible d'obtenir des valeurs d'imputation statique (*cold deck*) appropriées, c'est-à-dire des valeurs initiales pour ces combinaisons de valeurs « inconnues » pour une donnée variable puisqu'elles n'existent pas dans le monde « réel ».

175. La solution au problème présenté ci-dessus alourdit le travail de traitement des données, mais aboutit à un résultat moins entaché d'erreurs ou plus « pur ». Le programme de vérification contrôle tout d'abord la validité des codes attribués aux rubriques. Si l'enregistrement relatif à la personne considérée ne comporte pas un code valide attribué à la rubrique en question, alors la matrice d'imputation n'utilise pas ladite rubrique pour cet enregistrement. Les personnes chargées du traitement des données peuvent faciliter cette opération en mettant au point une matrice d'imputation simplifiée. Considérant à nouveau l'exemple précédent, au cas où le programme de vérification doit imputer une valeur à la variable « état matrimonial » (en raison de l'absence de cette donnée), la matrice d'imputation aura normalement les deux dimensions suivantes : âge et lien avec le chef de ménage. Si, après vérification, le programme établit qu'aucun code valide n'est attribué à la variable âge, il impute alors l'état matrimonial en fonction du seul lien avec le chef de ménage. Puisque la vérification concernant le lien avec le chef de ménage précède celle qui concerne l'état matrimonial, le code lien avec le chef de ménage sera valide. Le programme utilise ces mêmes principes pour toutes les procédures d'imputation dynamique.

#### e) Établissement de la matrice d'imputation

176. Il appartient aux spécialistes du sujet, en collaboration avec les informaticiens, d'établir les matrices d'imputation appropriées (certaines équipes de vérification utilisent plusieurs de ces matrices). Seules les réponses valides servent à la mise à jour des matrices, les équipes de vérification s'abstenant par ailleurs d'utiliser les valeurs déjà attribuées ou imputées. Il incombe aux spécialistes du sujet comme aux informaticiens de vérifier la cohérence et le caractère exhaustif des spécifications de vérification et des valeurs d'imputation dynamique (*hot deck*).

177. Il faut consacrer tout le temps nécessaire et réfléchir soigneusement à la mise au point d'une matrice d'imputation; il convient notamment d'étudier les pos-

<sup>6</sup> La meilleure pratique de vérification n'implique pas l'utilisation de valeurs vérifiées dans les données variables. Parfois, cette pratique est difficile à suivre, en raison du facteur temps pour les résultats ou de difficulté de programmation informatique. En pareil cas, l'une des variables serait imputée, et sa valeur placée dans les données variables appropriées, puis utilisée pour imputer les variables suivantes.

sibilités d'utilisation des enregistrements administratifs et des résultats des enquêtes ou recensements précédents. Même au terme de ce travail, les équipes de vérification ont intérêt à ne pas utiliser au hasard les matrices d'imputation. En l'absence de cohérence interne individuelle, l'harmonisation des matrices d'imputation exige un travail considérable. Par ailleurs, lorsque ces mêmes matrices n'utilisent pas des conventions normalisées, le personnel chargé de la vérification doit les examiner une à une.

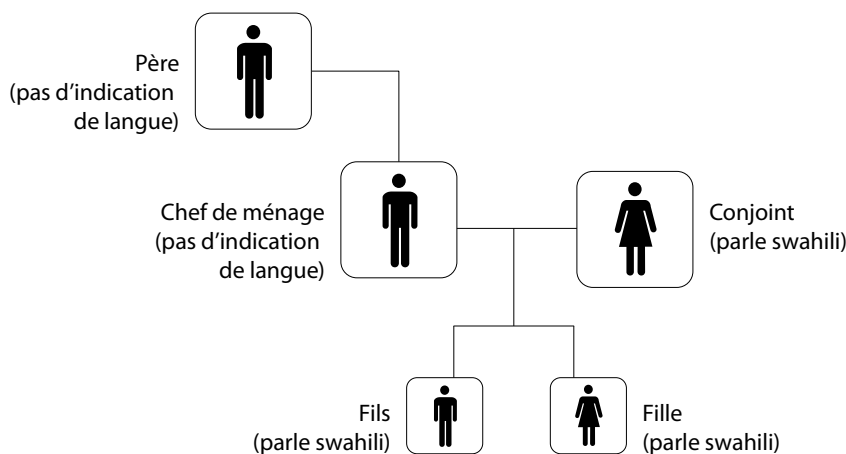
178. Bien que, pour les exemples cités dans le présent manuel, chaque case des matrices d'imputation contienne une valeur et une seule, certaines équipes de vérification conservent pour chacune plusieurs valeurs possibles, ce qui a pour effet d'introduire une dimension supplémentaire. On peut imaginer à cet égard une matrice à deux dimensions, avec une troisième dimension en profondeur. Si les âges de chacun des enfants d'une famille sont inconnus, par exemple, dans une famille comptant quatre enfants de sexe masculin, l'ordinateur n'attribuera pas la même valeur quatre fois et s'abstiendra donc de « créer » des quadruplés; il attribuera en revanche quatre âges différents. Toutefois, même dans ce cas, la même valeur peut être attribuée à plusieurs reprises, selon le contenu des matrices d'imputation.

#### f) Matrices d'imputation normalisées

179. L'utilisation de matrices d'imputation normalisées permet de simplifier le processus de vérification. Il est possible de mettre à l'essai et d'appliquer rapidement des matrices d'imputation de dimensions normalisées et se reportant à différentes variables sociales et économiques, telles que le groupe d'âge et le sexe.

180. Par exemple, le bureau national de recensement/statistique peut souhaiter créer une matrice d'imputation permettant d'attribuer le code correspondant à la langue lorsque aucun code n'est indiqué. Le programme de vérification devra, de façon pratiquement certaine, chercher en premier lieu, au sein du ménage considéré, la présence d'une autre personne signalée en tant que locuteur d'une langue donnée. Sinon, le programme peut choisir la langue d'une personne dont les données ont été précédemment vérifiées, du même sexe et du même groupe d'âge (après mise à jour de la matrice d'imputation en cas de validité des trois rubriques). Cette procédure permettra d'obtenir une valeur probable du code langue, puisque les personnes qui par-

Figure 14  
Exemple de défaut d'indication de la langue du chef de ménage  
et du père du chef de ménage





lent la même langue ou des langues voisines se trouvent généralement dans des zones géographiquement proches.

181. Dans le schéma de la figure 14, aucune valeur n'est attribuée à la variable « langue » concernant le chef de ménage. Pour une raison quelconque, le lecteur optique ou l'opérateur n'a sans doute pas saisi l'indication ou le code de la langue, à moins qu'il n'y ait eu un autre problème. Or, puisque le conjoint et les enfants parlent tous swahili, cette langue peut être imputée au chef de ménage, ainsi qu'au père du chef de ménage dont la langue n'a pas non plus été indiquée. On remarquera que le chef de ménage du schéma de la figure 14 est une femme.

182. Lorsqu'il n'y a d'indication de langue pour aucune des personnes du ménage, le programme de vérification doit procéder autrement. Il examine les autres variables susceptibles de fournir une estimation indirecte de la langue utilisée. Parfois, la race, le groupe ethnique ou le lieu de naissance peuvent donner une indication de la langue à imputer; en présence d'un indicateur de ce type, l'équipe de vérification peut alors décider de l'utiliser pour déterminer la langue du ménage. Sinon, le programme peut se servir des données d'âge et de sexe. Dans ce cas, la matrice d'imputation pourrait se présenter tel qu'indiqué à la figure 15.

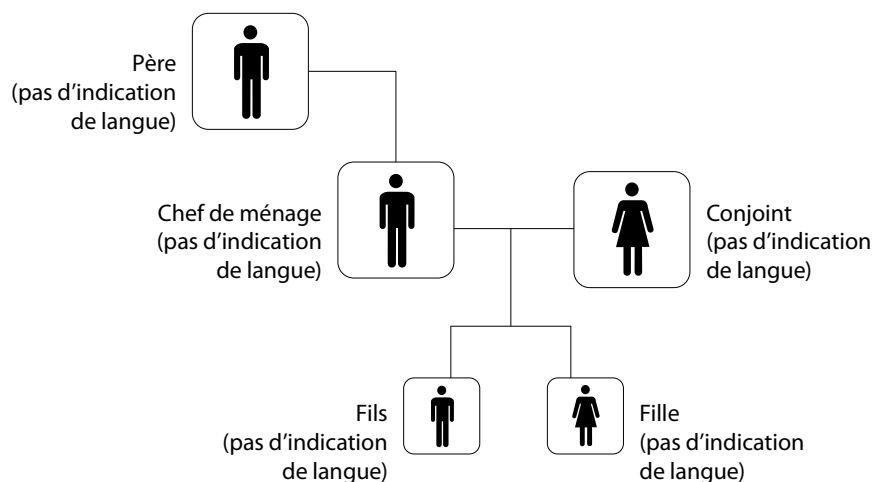
Figure 15  
Valeurs initiales d'une matrice d'imputation dynamique de la langue

Sexe	Âge					
	Moins de 15 ans	15-29 ans	30-44 ans	45-59 ans	60-74 ans	75 ans et plus
Homme	Langue 1	Langue 1	Langue 1	Langue 1	Langue 1	Langue 2
Femme	Langue 1	Langue 1	Langue 1	Langue 1	Langue 1	Langue 2

183. Si l'on décide d'imputer une valeur, le programme attribue au chef de ménage un code de langue en fonction du groupe d'âge et du sexe. Dans ce cas, les données dans la matrice d'imputation proviendront uniquement des précédents chefs de ménage puisque toutes les autres personnes d'un ménage donné se verront attribuer le même code de langue que le chef de ménage.

184. À ce stade, si à l'intérieur du ménage, personne ne déclare parler une langue particulière, le programme de vérification utilise la matrice d'imputation pour

Figure 16  
Exemple de membres d'un ménage dans lequel aucune langue n'a été attribuée



attribuer une langue au chef de ménage en s'appuyant sur l'âge et le sexe de ce dernier. La langue ainsi attribuée est celle qui a été identifiée le plus récemment dans le fichier de données comme étant parlée par un autre chef de ménage du même âge et du même sexe. Puisque la matrice d'imputation est « mise à jour » de façon continue au fur et à mesure de l'identification de données acceptables, la langue attribuée est susceptible d'être une langue parlée par l'ensemble de la communauté considérée (voir figure 16).

185. Dès le début du processus de vérification, les règles fixées donneront lieu à des exceptions. Ainsi, il faut veiller à prendre soin des éventuelles différences de langue lorsqu'on passe d'une zone géographique à une autre. Dans certains pays, il faut par ailleurs prêter attention aux combinaisons locales de locuteurs. Toutefois, même dans ce cas, à moins que certaines langues ne soient sous-déclarées de façon sélective, le processus d'imputation devrait conduire à des pourcentages sensiblement identiques de valeurs attribuées et non attribuées.

186. Une autre vérification pourrait porter sur la religion. Là encore, l'imputation de réponses concernant la variable religion est possible à partir des données d'âge et de sexe. Le programme de vérification poursuivra la mise à jour lorsque toutes les informations requises seront disponibles et utilisera les réponses fournies par la matrice d'imputation pour les affecter aux données « inconnues ». Cette matrice d'imputation présentera un aspect voisin de celle obtenue pour la variable langue, les données concernant la religion étant substituées à celles concernant la langue dans les cases correspondantes.

187. Cette description suppose une approche séquentielle de type descendant. Avec les méthodes sophistiquées, comme celle de Fellegi-Holt et la nouvelle méthodologie d'imputation (NIM, *New Imputation Methodology*) [voir annexe V], les équipes de vérification réalisent simultanément tous les contrôles connexes. La procédure décrite suppose par contre l'existence d'un ordre approprié de réalisation des vérifications.

188. Nombre de caractéristiques économiques telles que l'appartenance à la population active, le nombre d'heures de travail au cours de la semaine passée et le nombre de jours de travail au cours de l'année écoulée, peuvent être imputées sur la base de caractéristiques similaires. Grâce à l'utilisation de matrices d'imputation similaires, le programme de vérification peut contrôler rapidement la valeur des caractéristiques économiques utilisées et l'ensemble du processus devrait s'accélérer.

189. Il est parfois difficile d'obtenir des données dûment vérifiées pour les premières matrices d'imputation. D'ordinaire, la prise en compte de données non vérifiées dans une matrice d'imputation est a priori écartée; ainsi, le programme ne doit pas utiliser de données de sexe ou d'âge pour les affecter à la matrice d'imputation avant de les avoir contrôlées. Aussi les toutes premières matrices d'imputation utiliseront-elles des variables différentes n'exigeant pas de vérification ou des variables dont la valeur ne peut être modifiée. Quant à la première matrice d'imputation de données démographiques, le processus pourrait utiliser le nombre de personnes de l'unité d'habitation, celui-ci étant égal à 0 dans le cas des unités inoccupées.

190. Pour les vérifications des données sur l'habitat en règle générale, il serait également possible d'utiliser le nombre de personnes présentes dans les unités d'habitation en tant que dimension initiale de la première matrice d'imputation; l'équipe de vérification pourrait cependant modifier les ajustements concernant les unités d'habitation de manière à tenir compte des unités inoccupées. Par exemple, si la première vérification des données sur l'habitation porte sur le « matériau de construction des murs extérieurs » ou sur le « type de murs », les valeurs initiales pourraient être choi-

sies d'après le nombre de personnes présentes dans l'unité d'habitation, notamment une valeur particulière correspondant au cas d'une unité d'habitation inoccupée.

191. Lorsque l'unité d'habitation est inoccupée mais la donnée « type de murs » est valide, le programme de vérification met à jour la première case avec la valeur attribuée au type de murs extérieurs. Si le type de murs est connu, pour une unité d'habitation inoccupée, le programme met à jour la case qui correspond au nombre de personnes qui occupent ladite unité de logement. Si le matériau de construction des murs extérieurs est inconnu, la matrice d'imputation attribuera toutefois une valeur à la donnée « matériau de construction des murs extérieurs » en fonction du nombre de personnes de l'unité d'habitation.

192. Après avoir utilisé une première fois cette matrice d'imputation, l'équipe de vérification peut ensuite souhaiter passer à une autre caractéristique de l'habitat, telle que le « type de toiture » ou les « modalités de jouissance ». Quelle que soit la caractéristique choisie, elle doit être propre à chaque unité d'habitation et comporter une diversité suffisante pour que le même attribut ne soit pas retenu systématiquement. En effet, la sélection récurrente du même attribut risque de conduire à un type d'imputation pratiquement statique et non à une imputation dynamique (méthode *hot deck*). L'utilisation de la méthode dite d'imputation dynamique, par exemple dans le cas d'un logement collectif » constitué de casernes militaires, risque d'entraîner l'utilisation répétée de la même valeur, si l'âge et le sexe sont les seules caractéristiques retenues. En l'occurrence, tous les résidents seraient vraisemblablement du sexe masculin et se situeraient pour la plupart dans une tranche d'âge bien définie. Par conséquent, l'emploi de cette matrice particulière risque d'altérer la qualité des résultats. Lorsque la variable « modalité de jouissance » présente une diversité suffisante, avec des pourcentages suffisants de propriétaires et de locataires, son utilisation peut s'avérer efficace. Sinon, le choix pourrait porter sur la variable « type de toiture ».

193. En règle générale, nombre d'équipes de vérification estiment que l'utilisation de matrices d'imputation de dimensions comparables permet de réduire les contrôles, d'obtenir plus rapidement les résultats et sans doute d'améliorer leur précision.

**g) Cas de non-utilisation de la méthode d'imputation dynamique**

194. Si l'équipe de vérification décide de ne pas du tout utiliser la méthode d'imputation dynamique, l'ordre de réalisation des vérifications conserve son importance. Par exemple, l'âge est lié à de nombreuses rubriques, notamment au lien avec le chef de ménage, au niveau d'instruction, à l'emploi et à la fécondité (pour les femmes). Considérons les membres du ménage décrits à la figure 17 :

Figure 17  
Exemple de données concernant un chef de ménage et un enfant avec omission de l'âge de l'enfant

Personne	Lien avec le chef de ménage	Âge	Niveau de scolarité	Emploi	Profession	Nombre d'enfants
1	1	40	12	1	33	BLANC
3	3	X	7	BLANC	BLANC	BLANC

NOTE : X = Âge manquant  
BLANC = Sans objet

195. Les données concernant la personne n° 3 indiquent le code 3 de lien avec le chef de ménage (enfant) mais ne mentionnent pas d'âge. Afin de déterminer l'âge, le programme de vérification peut utiliser la différence d'âge entre le chef de ménage et

l'enfant (valeur obtenue par imputation statique ou méthode *cold deck* ou soit d'après une unité précédemment vérifiée). Si cette différence est égale à 25, par exemple, l'âge de l'enfant est égal à 15 ans (âge du chef de ménage, soit 40 ans, moins différence d'âge de 25).

196. Le nombre d'années de scolarité est également connu et égal dans ce cas à 7 ans. L'âge de 15 ans est en effet parfaitement susceptible de correspondre à ce niveau d'instruction. Puisque la fourchette des valeurs appropriées du nombre d'années de scolarité correspondant à un âge donné est moins étendue que la fourchette des âges correspondant à une différence d'âge donnée entre chef de ménage et enfant, il est préférable de vérifier en premier lieu si le nombre d'années de scolarité est adéquat. Si ce nombre d'années est mentionné, une différence d'âge obtenue par imputation statique (*cold deck*) ou dynamique (*hot deck*) peut servir à définir un âge approprié. Si le nombre d'années n'est pas connu, alors la différence d'âge entre chef de ménage et enfant peut servir à attribuer l'âge.

197. Or, même la différence d'âge peut figurer parmi les informations manquantes. De fait, dans la plupart des pays, le niveau de formation risque davantage d'être omis que l'âge. L'exemple ci-dessous décrit les différentes étapes du processus que peut suivre l'équipe de vérification en cas d'omission de l'âge et du niveau d'instruction.

Figure 18

**Exemple de données concernant un chef de ménage et un enfant avec omission de l'âge et du niveau de scolarité de l'enfant**

Personne	Lien avec le chef de ménage	Âge	Niveau de scolarité	Emploi	Profession	Nombre d'enfants
1	1	40	12	1	33	BLANC
3	3	X	X	BLANC	BLANC	BLANC

198. Les données de la figure 18 ne comportent aucune information d'âge ni de niveau d'instruction, mais contiennent par contre d'autres informations. La personne n° 3 n'est pas suffisamment âgée pour avoir eu des enfants (ou cette personne est de sexe masculin). D'après les indications disponibles sur l'emploi, une série de valeurs obtenues par imputation statique permettent de déterminer un âge, mais celui-ci sera inférieur à l'âge minimal acceptable pour occuper un emploi. Sinon, l'équipe de vérification procède à une imputation dynamique, la valeur de la matrice d'imputation permettant d'obtenir une valeur de l'âge. La valeur retenue sera vraisemblablement obtenue d'après l'âge du chef de ménage, en tant que variable permettant de maintenir la cohérence des données. Par exemple, si le chef de ménage est âgé de 20 ans et non de 40 ans, il ne conviendrait évidemment pas d'attribuer un âge de 14 ans à la personne n° 3. Lorsque l'âge est fixé, alors il est possible de déterminer également le niveau d'instruction, ce dernier devant être en effet cohérent avec l'âge et la situation du point de vue de l'emploi.

199. Si l'équipe de vérification décide d'imputer toutes les rubriques considérées ou presque, elle doit alors définir une stratégie d'élaboration logique du processus de vérification. En ce qui concerne les données démographiques, la vérification devrait commencer par tous les éléments comportant potentiellement des valeurs connues. Les équipes de vérification doivent mettre à profit les informations provenant d'enquêtes et d'archives administratives, de recensements précédents, de l'étude pilote effectuée pour le recensement considéré et d'autres informations disponibles pour faciliter

la prise en compte de chaque élément d'information dans la première matrice d'imputation et les suivantes. La définition des caractéristiques détaillées des matrices d'imputation est certes particulièrement liée au pays considéré, mais tous les bureaux de recensement/statistique disposent vraisemblablement à cet effet de certaines informations. L'expérimentation de divers ensembles de variables dans les matrices d'imputation dynamique permettrait au pays de déterminer l'ensemble le plus approprié.

200. Nombre de progiciels de vérification conservent au fur et à mesure de leur déroulement la valeur du nombre de personnes de l'unité d'habitation. Une matrice d'imputation pour la détermination des valeurs inconnues de la variable « sexe » par exemple, permet d'imputer la valeur de cette rubrique d'après le nombre d'occupants de l'unité de logement. Par conséquent, la valeur à choisir pour une personne de sexe non spécifié ou non valide dans le cas d'un ménage constitué d'une personne unique, pourrait être « sexe masculin ». Enfin, pour un ménage de trois personnes, la valeur pourrait être « sexe masculin », etc. La matrice d'imputation ne serait utilisée qu'en dernier ressort, une fois terminés tous les contrôles et les vérifications de cohérence, portant notamment sur le sexe du chef de ménage et du conjoint et la présence de données sur la fécondité.

#### h) Quelle doit être la taille des matrices d'imputation ?

201. La plupart des progiciels informatiques peuvent traiter des matrices d'imputation multidimensionnelles. Les observations suivantes doivent être prises en compte lors de la création des matrices d'imputation.

##### i) *Problèmes posés par une taille excessive de la matrice d'imputation*

202. Pour certains bureaux nationaux de recensement/statistique, l'excès de perfectionnisme des vérificateurs est l'un des principaux problèmes rencontrés dans le cadre de la collaboration entre spécialistes du sujet et informaticiens. Les concepteurs des progiciels de vérification se laissent volontiers emporter par leur inspiration, de telle sorte que la programmation prend beaucoup plus de temps que nécessaire et ralentit le dépouillement du recensement et de l'enquête. L'équipe de vérification doit décider par exemple, en vue de la détermination de l'âge, qu'il faut également tenir compte du « nombre d'enfants » dans le cas des individus de sexe féminin, outre les variables « sexe », « niveau d'instruction » et « appartenance à la population active ». Le fait d'ajouter la variable « nombre d'enfants » est susceptible d'améliorer légèrement l'estimation de l'âge, bien que cette amélioration ne justifie pas nécessairement la complexité accrue de la programmation. Il incombe donc aux équipes de vérification de décider des dimensions de la matrice d'imputation qui fournira les meilleurs résultats, tant en termes de précision que d'efficacité. Les matrices d'imputation de taille excessive (comportant un trop grand nombre de cases) ne peuvent être parfaitement actualisées, auquel cas, des valeurs obtenues par imputation statique risquent d'être utilisées à tort à leur place.

##### ii) *Connaissance précise des tâches effectuées au moyen de la matrice d'imputation*

203. Outre la taille excessive des matrices d'imputation, les chemins de vérification risquent d'être compliqués. Il importe de veiller à ce que les spécialistes du sujet et les informaticiens soient en mesure de suivre l'exécution de toutes les séquences. Ils doivent conjointement s'assurer du fait que la matrice d'imputation exécute la tâche voulue. Là encore, la collaboration entre spécialistes et informaticiens est indispensable afin de vérifier la bonne mise en œuvre de chaque variable ou de chaque dimension

de la matrice d'imputation. De plus, ils doivent vérifier que toutes les combinaisons fonctionnent correctement.

iii) *Problèmes posés par une matrice d'imputation de taille insuffisante*

204. La matrice d'imputation est trop petite si les variables sont trop peu nombreuses ou si, en raison des regroupements pratiqués (groupes d'âge ou niveaux de formation en nombre insuffisant), la même valeur est imputée à plusieurs reprises avant d'être mise à jour. Par exemple, sans prise en compte de la variable sexe, dans un tableau d'âges, tous les enfants d'une famille sont davantage susceptibles de se voir attribuer le même âge lorsque cette donnée est inconnue. Avec le concours des informaticiens, les spécialistes du sujet devraient s'employer à tester les matrices d'imputation pour toutes les combinaisons possibles et veiller à ce qu'aucune combinaison n'intervienne trop fréquemment.

iv) *Rubriques difficiles à corriger aux moyen de matrices d'imputation*

205. Certaines rubriques, telles que la « profession » ou le « secteur d'activité » se sont avérées manifestement difficiles à corriger. Tandis que l'utilisation de matrices d'imputation distinctes pour la profession et le secteur d'activité risquent d'aboutir à des résultats incohérents, le coût et la difficulté d'un travail systématique de vérification de tous les couples de profession et de secteur d'activité peuvent être considérables. Ainsi, la détection de coiffeurs qui travaillent dans des usines de transformation du poisson doit donner lieu à une forme ou une autre de vérification. De plus, la multiplicité des professions et des secteurs d'activité risque de compliquer énormément le processus d'imputation dynamique. De fait, pour certaines rubriques, l'équipe de vérification conviendra vraisemblablement de l'inefficacité des contrôles et choisira de leur substituer l'utilisation des réponses « non indiqué » ou « non signalé ». Sinon, le recours à une méthode d'imputation statique (*cold deck*) pourra suffire.

#### 4. Contrôle des matrices d'imputation

206. La matrice d'imputation utilisée dans un progiciel de vérification devrait présenter une structure de base proche de celle de la figure 19. Les spécifications de vérification doivent déterminer les tableaux de valeurs utilisées pour les imputations et déterminer le jeu de valeurs initiales obtenues par imputation statique.

a) *Constitution de la matrice statique initiale*

207. La procédure décrite ci-dessous met à jour la matrice d'imputation chaque fois qu'elle trouve une personne à laquelle sont associées des valeurs valides pour chacune des trois rubriques considérées, en l'occurrence le « lien avec le chef de ménage », le « sexe » et « l'âge ». Toutefois, lorsque le programme de vérification trouve une valeur non valide (ou absente) de la variable « sexe », la matrice d'imputation choisit une valeur d'après les codes valides attribués au lien avec le chef de ménage et au sexe (variables préalablement vérifiées).

Figure 19  
Exemple de jeu de valeurs utilisées dans un tableau d'imputation statique et codes imputés correspondants

Valeurs initiales pour la matrice d'imputation 22 A01-AGE-FM-SEXRL (2,6)

Chef de ménage	Conjoint	Enfant	Autre parent	Parent	Non indiqué	Sexe
40	40	10	20	65	20	Homme
40	40	10	20	65	20	Femme

```

.
.
.
si ÂGE = 0:98
  A01- ÂGE - FM - SEXRL (SEXE, RELATION) = ÂGE
Autrement
  message « Âge est inconnu, donc attribué » ÂGE
  écrire « âge inconnu donc attribué, Âge = » ÂGE
  ÂGE attribué = A01 - ÂGE - FM - SEXRL (SEXE, RELATION)
  message « ÂGE est connu maintenant » ÂGE
Terminé - si
.
.
.

```

b) Messages relatifs aux erreurs

208. Les progiciels de vérification doivent proposer plusieurs moyens pour garantir l'exécution effective et correcte des vérifications et des imputations. Deux de ces moyens, les instructions de message et les instructions d'écriture, sont passés en revue ci-dessous.

209. L'affichage d'un message, tel qu'indiqué ci-dessus à la figure 19, est une source d'information. Cette instruction génère des messages et des récapitulatifs particuliers (nombre total d'occurrences du message) pour divers niveaux d'agrégation géographique (par exemple, secteur de dénombrement, circonscription administrative secondaire, circonscription administrative principale), comme pour chaque questionnaire. Pour l'ensemble des questionnaires, un rapport récapitulatif peut se présenter comme indiqué à la figure 20 ci-dessous :

Figure 20  
Exemple de rapport indiquant le nombre d'imputations par erreur

Décompte	Numéro d'erreur	Message	Numéro de ligne
-	14-1	Nombre excessif d'enfants par femme	2629
-	14-2	Nombre excessif d'enfants par femme	2645
2	14-3	Non-indication de jeunes gens présents	2669
2	14-4	Non-indication de jeunes filles présentes	2678
33	14-5	Non-indication du mois de la dernière naissance	2723
7	15-6	Aucun nombre d'enfants; différence d'âge correcte entre la mère et l'enfant	2892

**Note :** Dans ce cas, la mention « 14 » désigne simplement la rubrique n° 14 d'une série de données; les erreurs sont numérotées séquentiellement.

210. Un rapport établi pour chaque questionnaire particulier (figure 21) indiquera le cas échéant le numéro du questionnaire, ainsi que tous les codes géographiques spécifiés. Il pourrait également indiquer la liste des erreurs détectées par le programme, classées par rubrique (dans ce cas, l'âge) et par numéro de ligne du logiciel, tel qu'indiqué ci-dessous dans la colonne de droite. Dans cet exemple, l'âge était laissé

en blanc, mais la matrice d'imputation a permis de calculer un âge de 48 ans, d'après les valeurs du lien avec le chef de ménage et du sexe concernant cette personne. En l'occurrence, l'âge en question n'était pas spécifié, mais l'instruction d'édition d'un message pourrait également mentionner cette information si nécessaire.

Figure 21

**Exemple de rapport d'erreurs contenues dans un questionnaire**

	Questionnaire ID : 01 01 017	Numéro de ligne
AGE (1) =	Âge inconnu, donc imputation	#46
AGE (1) = 48	Âge maintenant connu	

211. Bien entendu, s'il paraît logique d'énumérer toutes les erreurs relevées sur des tests d'échantillonnage ou sur de petits ensembles de données sélectionnés, l'ampleur des résultats serait excessive et peu commode (et perdrait à la longue tout son sens). Pour éviter cet écueil, il conviendrait de mettre en place un mécanisme permettant de bloquer la totalité ou une partie des problèmes posés par certains des questionnaires utilisés pour le recensement.

## c) Établissement à la demande de listes d'erreurs

212. Le logiciel peut également proposer une autre instruction permettant une analyse plus détaillée des spécifications de vérification et du déroulement des opérations correspondantes. Cette instruction peut servir à présenter les données avant modification, puis toutes les modifications introduites. De cette façon, l'analyste peut acquérir la certitude du bon déroulement des vérifications selon tous les chemins définis. Les résultats obtenus peuvent se présenter tel qu'indiqué à la figure 22. La première ligne indique les variables (par exemple, province, lien avec le chef de ménage, sexe, âge). Ensuite, les données d'entrée sont indiquées, puis l'erreur détectée (en l'occurrence, pas d'indication d'âge) et enfin, les données obtenues après modification.

Figure 22

**Exemple de liste imprimée d'erreurs supplémentaires pour questionnaire portant sur plusieurs variables**

	Province	District	Chef de ménage	Lien avec le chef de ménage	Sexe	Âge
Données d'entrée	01	01	17	1	1	
Erreur	Âge inconnu, donc imputation âge = BLANC					
Données vérifiées	01	01	17	1	1	48

213. Cette procédure aide l'équipe de vérification à déterminer si le programme suit les chemins appropriés.

214. Le contrôle est une partie importante du travail de vérification des recensements et des enquêtes. La méthode décrite ci-après est un moyen de contrôle des procédures de vérification. La marche à suivre pourrait consister à demander aux spécialistes de commencer systématiquement par la définition d'un ménage « parfait ». On entend par « ménage parfait » un ménage complet : chef de ménage, conjoint, enfant, autres parents et non-parents, avec indication de toutes leurs caractéristiques. Le ménage parfait doit satisfaire à toutes les vérifications sans donner lieu à aucune erreur. La procédure se poursuit tel qu'indiqué ci-dessous :



- a) Le personnel chargé de traiter les données introduit une erreur unique dans chaque ménage, de façon séquentielle, correspondant à la séquence des spécifications de vérification et au déroulement du programme de contrôle;
- b) L'analyste contrôle ensuite tous les chemins au début du processus de vérification;
- c) Lorsque la vérification suit correctement tous les chemins, le personnel chargé de traiter les données applique le programme à un échantillon de jeu de données complet, y recherche les anomalies et effectue les modifications éventuellement nécessaires;
- d) Enfin, le personnel chargé du traitement exécute le programme sur la totalité du jeu de données.

215. Après avoir constaté le fonctionnement satisfaisant des messages et avoir apporté les modifications appropriées, le personnel informaticien peut décider de les désactiver au niveau du questionnaire. Si les pays de taille importante devaient traiter la totalité de leurs jeux de données en maintenant la publication des messages correspondant à chaque questionnaire, il en résulterait une quantité prohibitive de lignes d'impression et de papier. Toutefois, le rapport récapitulatif concernant ces messages doit être conservé, puisqu'il donne des indications intéressantes aux différents niveaux d'agrégation géographique. L'état imprimé ressemblera à celui de la figure 22.

216. Les vérifications automatisées comprennent généralement une procédure de sauvegarde. L'analyse rétrospective fait apparaître toutes les modifications apportées aux données ainsi que des repères indiquant lesdites modifications et les valeurs substituées. La consultation de l'analyse rétrospective déterminera si les changements apportés sont suffisamment peu nombreux pour autoriser l'acceptation du groupe d'enregistrements.

217. Si une rubrique particulière comporte un nombre d'erreurs excessif, il est possible qu'elle n'ait pas fait l'objet de tests préalables adéquats, soit isolément, soit en rapport avec les autres rubriques. Parfois, les agents de recensement sont induits en erreur et recueillent les informations sur la fécondité uniquement auprès des adultes de sexe masculin et non auprès des femmes. Si les données sont collectées systématiquement de cette façon, l'équipe de vérification devra sans doute demander aux programmeurs de transférer les données sur la fécondité des hommes aux femmes, à l'intérieur d'un couple marié. Sinon, elle ne peut pas faire grand-chose à ce stade pour corriger cette erreur.

218. Généralement, le programme de vérification doit être appliqué à plusieurs fichiers différents pour couvrir toutes les situations. En outre, le personnel informaticien devra effectuer des changements en raison des erreurs de syntaxe ou de logique. Même les informaticiens les plus expérimentés peuvent taper accidentellement un signe « plus grand que » au lieu d'un signe « plus petit que » et l'erreur n'est ensuite constatée qu'après plusieurs passages, puisque ce type de problème n'est pas toujours immédiatement apparent. De même, de petites erreurs dans le schéma logique de traitement risquent de ne pas apparaître d'emblée. Là encore, spécialistes du sujet et informaticiens doivent travailler de concert à la résolution de ces problèmes, si possible dès le début du processus de vérification.

#### d) Nombre de passages du programme de vérification

219. Comme on l'a indiqué, dès que le questionnaire est prêt, il convient de commencer à définir et tester les spécifications et programmes de vérification. Les diverses rubriques doivent être conçues séparément si l'on utilise une approche de

type descendant, mais mêmes si plusieurs variables doivent être vérifiées en même temps, les vérifications des différentes rubriques doivent être contrôlées sur de petits fragments du jeu de données complet. Les spécifications de vérification doivent être conçues par les spécialistes du sujet et les divers programmes de vérification doivent ensuite être appliqués par les programmeurs. Les programmeurs peuvent ensuite définir la vérification dans son ensemble et l'appliquer à des fragments de plus en plus grands du jeu de données, tout en l'améliorant en cours de route.

220. En règle générale, aussi bien pour certaines parties du programme de vérification que pour le programme dans son ensemble, il y a tout intérêt à appliquer trois fois ce programme tel qu'indiqué ci-dessous :

221. Le premier passage affecte aux matrices d'imputation des valeurs réelles et non les valeurs créées dans la matrice statique initiale. Certains pays utilisent des données d'une autre provenance — enquêtes ou recensements antérieurs ou données administratives — pour déterminer les valeurs à faire figurer dans un tableau d'imputation statique. Le personnel informaticien vérifie le jeu complet de données ou une grande partie de celui-ci pour obtenir les valeurs à faire figurer dans la matrice d'imputation. Les données d'imputation statique tirées du jeu de données réelles ont plus de chances d'être exactes et actualisées. Les vérifications utilisent seulement 2 % du contenu de cette matrice statique initiale : le reste est constitué de valeurs d'imputation dynamique.

222. Le deuxième passage du programme procède à la vérification proprement dite et comporte plusieurs passages répétés afin de couvrir toutes les situations. À ce stade, le personnel informaticien devra introduire des modifications pour corriger les erreurs dues aux défauts de syntaxe ou de logique. Or, l'informaticien le plus expérimenté étant faillible et tel problème particulier ne se manifestant pas forcément immédiatement, il faut souvent plusieurs passages avant de constater les erreurs de ce type. De même, les petits défauts du schéma logique de traitement ne sont pas toujours apparents d'emblée.

223. Le troisième passage établit de façon certaine : 1) que le jeu de données ne comporte plus d'erreurs; et 2) que le programme de vérification n'en a pas introduit de nouvelles. Pour ce dernier passage, les états imprimés ne doivent plus faire apparaître aucune erreur. Si des erreurs persistent, le schéma logique de traitement est probablement défectueux et le personnel informaticien doit alors le modifier. De plus, ce passage indique généralement si la vérification a introduit accidentellement de nouvelles erreurs imputables au schéma logique de traitement.

## **5. Indicateurs d'imputation**

224. L'utilisation d'indicateurs d'imputation permet de conserver l'information relative à des données non vérifiées. Tel qu'indiqué plus haut, nombre d'équipes de vérification sont préoccupées par la perte d'information potentielle en cas de modification des réponses non vérifiées. Lorsqu'une valeur est changée en raison d'une incohérence, les équipes de vérification souhaitent parfois conserver la ou les valeurs initiales afin de procéder à des études démographiques ou des analyses d'erreurs plus poussées, une fois le recensement terminé. Ainsi, spécialistes du sujet et programmeurs souhaiteront analyser différents aspects des données omises, non valides ou incohérentes. Les membres de l'équipe de vérification devront s'assurer de la cohérence et de la compatibilité des distributions imputées et non imputées, afin de déceler l'apparition éventuelle d'une erreur systématique dans le programme de vérification et d'imputation. Par exemple, les informaticiens peuvent utiliser incidemment uni-

quement les valeurs d'imputation statique parce que le programme a omis de mettre à jour la matrice d'imputation. En cas de réalisation d'un essai précensitaire national, l'équipe de vérification doit parfois étudier les liens entre certaines variables à la suite de l'essai pour mettre au point la version définitive du questionnaire. Lors des recensements précédents, alors que les micro-ordinateurs équipés de disques durs de grande capacité n'étaient pas encore répandus, nombre de bureaux de statistique n'avaient pas assez de place sur leurs bandes magnétiques ou sur leurs supports de stockage pour conserver une quantité supplémentaire de données; en revanche, de nos jours et dans la plupart des pays, la conservation des informations relatives à des données non vérifiées n'est plus problématique.

225. Certains pays décident de maintenir une simple variable comptable binaire en tant qu'indicateur pour chaque rubrique. Cette méthode est simple et n'utilise qu'un octet pour chaque variable. Par exemple, le United States Census Bureau place des indicateurs d'imputation pour chaque variable à la fin de chaque enregistrement, aussi bien pour les données de population que pour les données d'habitat. Ainsi, pour chaque variable concernant l'habitation, l'indicateur a initialement la valeur 0, mais passe à la valeur 1 en cas de modification quelconque de la donnée élémentaire d'origine. Le programme utilisé ne conserve pas la valeur d'origine, quoique certains bureaux de statistique procèdent à la compilation desdites valeurs soit pour chaque enregistrement, soit au niveau de l'agrégat.

226. Il existe d'autres méthodes de sauvegarde des réponses non vérifiées. Dans l'exemple de la figure 23, le bureau national de recensement/statistique a modifié l'âge d'un conjoint en le faisant passer de 70 à 40 ans au moyen d'une matrice d'imputation. Il est alors facile de placer la valeur pré-imputation (en l'occurrence 70 ans) dans la zone réservée aux indicateurs d'imputation et de réserver la variable utilisée dans les tableaux destinés à la publication pour la valeur attribuée en définitive, en l'occurrence 40 ans. Afin d'examiner les changements introduits dans le jeu de données, le bureau de statistique peut établir des distributions de fréquence ou des tabulations croisées des valeurs allouées et non allouées. Si à la suite de cette analyse de l'incidence des vérifications sur le jeu de données, les tableaux obtenus au terme de la vérification semblent comporter des valeurs douteuses ou anormales, les équipes de vérification peuvent envisager une modification totale ou partielle du schéma de vérification; et vu l'augmentation observée ces dernières années de la capacité des disques durs, toutes les valeurs initiales peuvent être mises en mémoire sur les enregistrements en vue de leur utilisation ultérieure. Les bureaux de statistique tiendront probablement à conserver au moins deux fichiers car la tabulation d'un fichier ne contenant que les données vérifiées serait probablement plus rapide.

Figure 23  
**Exemple de données sur la population comportant des indicateurs associés aux valeurs imputées**

Personne	Sexe	Âge	Nombre d'enfants (NE)	Indicateur associé au sexe	Indicateur associé à l'âge	Indicateur associé au nombre d'enfants (NE)
1	1	40	BLANK			1
2	2	40	7		70	

227. La figure 24 se rapporte à une femme âgée de 13 ans dont les données indiquent qu'elle a un enfant (nombre d'enfants = 1). Or, l'équipe de vérification a convenu que l'âge minimal à la première naissance était de 14 ans et que la mention d'enfants

nés de femmes plus jeunes correspondait plus vraisemblablement à une erreur qu'à un fait. Comme toujours, cela soulève la question de savoir si ce type de cas représente un bruit dans le jeu de données ou bien une valeur réelle.

Figure 24

**Exemple d'indicateurs correspondant à une jeune femme dont l'enregistrement comporte des données sur la fécondité laissées en blanc, avec ajout d'un indicateur**

Personne	Sexe	Âge	Nombre d'enfants (NE)	Indicateur associé au sexe	Indicateur associé à l'âge	Indicateur associé au nombre d'enfants (NE)
<b>Données sur la fécondité laissées en blanc</b>						
4	2	13	1			
<b>Données sur la fécondité laissées en blanc avec indicateur</b>						
4	2	13	BLANC			1

228. Selon les règles de vérification, le processus d'imputation « efface » l'information concernant le nombre d'enfants. Signalons que l'indicateur relatif au nombre d'enfants est un peu plus complexe, puisqu'il doit tenir compte d'un vide (BLANC) qui a été imputé, ainsi que des valeurs numériques. Supposons que les démographes souhaitent étudier les effectifs et les caractéristiques des personnes de 13 ans dont les données indiquent qu'elles ont eu un enfant. Le personnel informaticien peut enregistrer l'information d'origine dans une zone de l'enregistrement réservée aux indicateurs, généralement à la fin. Ensuite, la série de tableaux publiés omettra l'information concernant le nombre d'enfants de cette femme, mais cette information restera disponible en vue de travaux de recherche ultérieurs. Par la suite, notamment lors de l'organisation d'une enquête de suivi, ou à l'occasion du recensement suivant, les équipes de vérification peuvent utiliser l'information concernant les enfants nés de femmes de 13 ans pour décider s'il faut abaisser l'âge limite de prise en compte.

229. Un des problèmes posés par l'utilisation des indicateurs d'imputation est la place considérable occupée dans l'ordinateur par la procédure décrite ci-dessus. Si les indicateurs répètent chaque variable, le jeu de données vérifiées occupera deux fois plus de place que le jeu de données non vérifiées. Pour de nombreux pays, cette exigence serait inacceptable dans la perspective d'un stockage à long terme. Les données d'origine et les vérifications pourraient cependant être enregistrées en vue de leur reconstitution ultérieure.

230. Les pays à très forte population préféreront utiliser des indicateurs d'imputation en procédant par échantillonnage, en vue de travaux de recherche. Par exemple, un pays peut souhaiter créer un jeu de données concernant une unité d'habitation sur 100. Le programme de vérification affectera alors des indicateurs sur ce jeu de données réduit, ce qui facilitera l'évaluation de l'incidence de la vérification sur la qualité des données, ainsi que la détermination des différences entre données vérifiées et non vérifiées.

## F. Autres systèmes de vérification

231. La plus grande partie du présent manuel s'emploie à décrire l'application des méthodes de type descendant à la vérification automatisée des recensements et des enquêtes. De fait, quelques pays mettent en œuvre à cet effet une procédure différente et plus compliquée connue sous le nom de vérification à plusieurs variables (voir section D.2 ci-dessus). Fellegi et Holt (1976) ont été les premiers à mettre au point ces

méthodes couramment appliquées aux principales variables des recensements ou des enquêtes : âge, sexe, lien avec le chef de ménage et état matrimonial. Ces procédures sont néanmoins applicables à tout un groupe de variables ou à l'ensemble des variables figurant sur le questionnaire d'un recensement ou d'une enquête. Selon cette méthode, le programme de vérification examine simultanément les réponses fournies à toutes ces rubriques pour une personne ou pour l'ensemble des personnes d'un ménage afin d'identifier les réponses manquantes ou incohérentes. Lorsque l'on constate la présence de données inconnues (blanc), non valides ou incohérentes, on procède à une série de tests pour déterminer laquelle de ces données est la plus erronée et l'on commence par modifier cette donnée. Puis on répète ces tests pour s'assurer qu'il ne reste plus de données non valides ou incohérentes; s'il en reste, on modifie la donnée qui pose encore le plus de problèmes. On répète cette procédure jusqu'à ce qu'il ne reste plus d'erreurs.

232. Statistique Canada a mis au point la méthode Fellegi-Holt et l'a utilisée pour les recensements au Canada de 1976 à 1991. Pour le recensement du Canada réalisé en 1996, cette approche a été perfectionnée et intitulée Nouvelle méthodologie d'imputation (NIM). Cette approche a permis pour la première fois de « résoudre des problèmes à grande échelle [de vérification et d'imputation], en procédant par imputation simultanée de variables numériques et qualitatives » (Bankier, Houle et Luc, n.d.).

233. Si le processus de vérification utilise la méthode traditionnelle d'imputation dynamique (*hot deck*), les informations d'imputation concernant une série de rubriques élémentaires du questionnaire peuvent provenir d'un grand nombre d'individus différents, selon l'information utilisée pour la mise à jour de la matrice d'imputation. Par exemple, si le sexe de la personne A, le lien avec le chef de ménage et l'état matrimonial sont corrects, ces valeurs serviront à la mise à jour des matrices appropriées. Par contre, si l'âge de la personne A est manquant ou non valide, il ne servira évidemment pas à mettre à jour les matrices d'imputation. En fait, les autres rubriques serviront à mettre à jour cette valeur. Si la valeur de la variable « sexe » de la personne suivante est incohérente, et en cas d'imputation du sexe, la personne A fera alors office de « donneur ». Si l'âge est également inconnu, le programme de vérification utilisera l'âge d'une autre personne.

234. La nouvelle méthodologie d'imputation utilise des donneurs pour les rubriques considérées, dans l'espoir que toutes les informations manquantes ou incohérentes pourront provenir d'un donneur unique ou d'un petit nombre de donneurs. L'obtention de la totalité ou de la plus grande partie de l'information auprès d'un donneur unique exige l'enregistrement des fichiers de données dans la mémoire de l'ordinateur. Alors, en cas de données non spécifiées ou non valides d'âge et de sexe, les mêmes variables enregistrées fournissent des valeurs pour ces deux rubriques.

235. La méthode automatisée d'imputation dynamique doit s'attacher aux objectifs suivants :

- a) Le ménage dont les données ont été imputées doit ressembler étroitement au ménage dont la vérification a mis en évidence des erreurs;
- b) Les données imputées concernant un ménage doivent provenir si possible d'un seul donneur, plutôt que de plusieurs. En outre, le ménage dont les données ont été imputées doit ressembler étroitement au donneur unique;
- c) Différentes imputations de qualité identique, fondées sur les donneurs disponibles, doivent avoir une probabilité voisine d'être retenues, afin d'éviter un accroissement disproportionné de la taille de groupes d'effectif réduit, dont le rôle est néanmoins important (Bankier, Houle, et Luc, n.d).

236. Ces objectifs sont atteints quand on identifie tout d'abord les ménages dûment vérifiés qui sont aussi proches que possible de ceux qui ont été rejetés par la véri-

fication. Autrement dit, les deux ménages doivent coïncider en ce qui concerne le plus grand nombre possible de variables qualitatives et présenter seulement des différences limitées en ce qui concerne les variables numériques. Les ménages qui possèdent ces caractéristiques sont appelés « voisins les plus proches ». L'étape suivante consiste à identifier, pour chacun des voisins les plus proches, les plus petits sous-ensembles de variables ne coïncidant pas (tant numériques que qualitatives) qui, en cas d'imputation, permettent au ménage de passer avec succès la vérification. On procède ensuite à la sélection aléatoire de l'une des opérations d'imputation qui permettent de passer avec succès la vérification et de regrouper les ménages rejetés et les ménages acceptés (Bankier, Houle, et Luc, n.d.).

237. Le présent chapitre a été consacré aux procédures générales de vérification et de tabulation. Le chapitre III traite des vérifications structurelles, qui portent sur la première opération de vérification automatisée, qui est aussi la plus importante car c'est elle qui permet de déterminer que chaque unité d'habitation est incluse, et n'est incluse qu'une fois, et qu'elle se trouve à la place appropriée dans la hiérarchie du pays.

## Chapitre III

# Vérifications structurelles

238. Les vérifications structurelles contrôlent la couverture des enregistrements réalisés et déterminent comment les divers enregistrements s'ajustent entre eux. Ces vérifications structurelles doivent assurer que : *a)* les enregistrements relatifs à tous les ménages et à tous les logements collectifs à l'intérieur d'un secteur de dénombrement sont présents et figurent dans le bon ordre; *b)* que toutes les unités d'habitation occupées comportent des enregistrements relatifs à des personnes, contrairement aux unités inoccupées; *c)* que les ménages ne comportent aucun enregistrement personnel en double ou manquant; et *d)* que les secteurs de dénombrement ne comportent aucun enregistrement d'habitation en double ou manquant. Les vérifications structurelles s'emploient donc à garantir le caractère globalement exhaustif des questionnaires.

239. Les vérifications structurelles particulières réalisées pour un recensement ou une enquête sont susceptibles de devoir évoluer au cours du temps compte tenu de la rapidité de changement des techniques appliquées à la détermination et à la correction des erreurs de structure. Aussi le présent chapitre examine-t-il la question plus générale de la validité des données et des relations entre elles, à l'intérieur d'un enregistrement et d'un enregistrement à l'autre. Les chapitres IV et V traitent de rubriques particulières concernant la population et l'habitation.

### A. Vérifications géographiques

#### 1. Emplacement des locaux d'habitation (localité) [P1H]

240. Une localité, d'après la publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* (Nations Unies, 2008, par. 2.78), est définie comme « un groupement de population distincte... dont les membres occupent des locaux à usage d'habitation voisins et qui a un nom ou un statut localement reconnu ». Des informations supplémentaires concernant l'emplacement des locaux d'habitation figurent dans la définition des termes « localité » et « population urbaine et rurale » aux paragraphes 2.78 à 2.88 des *Principes et recommandations*. Il est essentiel que les personnes responsables de la réalisation des recensements de l'habitation étudient ces indications en raison de l'importance particulière des concepts géographiques servant à décrire l'emplacement des locaux d'habitation dans le cadre des recensements de l'habitat, tant pour l'exécution du recensement que pour l'exploitation des résultats (Nations Unies, 2008, par. 2.455).

241. Lors de la vérification des données d'emplacement, les codes géographiques doivent être parfaitement exacts. Ainsi, l'obtention de codes complets et exacts relatifs aux différents niveaux du découpage géographique en vue du traitement des données figure-t-elle parmi les tâches les plus délicates du recensement. En cas d'erreur de code géographique, les opérateurs de saisie risquent d'attribuer l'unité ou les unités d'habitation à une autre région du pays. Il est extrêmement difficile de corriger ce type d'erreur.

## 2. Résidences urbaines et rurales (P1L)

242. La distinction traditionnelle entre les zones urbaines et rurales d'un pays repose sur l'hypothèse selon laquelle dans les zones urbaines quelle que soit la définition que l'on en donne, le niveau de vie y est différent et en général plus élevé que dans les zones rurales. Dans bien des pays industrialisés, cette distinction n'est plus très nette et la principale différence entre les conditions de vie des zones urbaines et des zones rurales tient avant tout au degré de concentration de la population. Des différences entre les modes de vie et les niveaux de vie urbains et ruraux dans les pays en développement sont encore sensibles mais ces pays ont désormais grand besoin eux aussi, en raison de l'urbanisation rapide qu'ils connaissent, de données relatives aux zones urbaines de différentes dimensions (Nations Unies, 2008, par. 2.82).

### Encadré 3

#### Principes directeurs concernant les vérifications structurelles

Les vérifications structurelles doivent assurer l'exécution des tâches suivantes :

- ☞ Vérifier que chaque lot de secteurs de dénombrement a reçu les codes géographiques appropriés (provinces, districts, secteurs de dénombrement, etc.) et que la désignation des lots suit la pratique courante;
- ☞ Vérifier que toutes les unités d'habitation sont prises en compte et que tous les ménages d'un secteur de dénombrement sont inscrits;
- ☞ Regrouper les ménages au sein de leurs secteurs de dénombrement respectifs et regrouper les secteurs de dénombrement au niveau d'agrégation géographique supérieur approprié;
- ☞ Faciliter le choix entre les pages concernant les personnes et les pages concernant le ménage, comprises ou non dans le questionnaire ou en dehors, en fonction de l'effectif de la population et de l'organisation du questionnaire;
- ☞ Attribuer chaque enregistrement individuel au type d'enregistrement valide approprié;
- ☞ Traiter séparément les enregistrements concernant les logements ou locaux collectifs à usage d'habitation et ceux concernant les unités d'habitation;
- ☞ Vérifier l'existence d'une correspondance entre les différents types d'enregistrement : par exemple, aucune personne n'habite les unités inoccupées; une personne au moins habite les unités dites occupées;
- ☞ Veiller à ce que le nombre d'enregistrements personnels concernant chaque ménage corresponde au nombre total de personnes du ménage indiqué dans l'enregistrement relatif à l'habitation;
- ☞ Veiller à la présence du nombre correct de questionnaires en cas d'utilisation de documents multiples pour un ménage donné et vérifier qu'ils sont liés entre eux de manière adéquate;
- ☞ Éliminer les enregistrements en double, aussi bien à l'intérieur des ménages (personnes figurant en double) que d'un ménage à l'autre (ménage inscrit en double ou partiellement inscrit en double), afin d'éviter les erreurs de surcouverture;
- ☞ Classer les enregistrements vides à l'intérieur d'un type d'enregistrement particulier;
- ☞ Traiter les unités d'habitation manquantes.



243. La plupart des pays définissent les zones géographiques dites « urbaines » et « rurales » préalablement au recensement et introduisent les ajustements nécessaires une fois les données de recensement recueillies. En cas d'attribution de codes correspondant à une résidence urbaine ou rurale (par exemple, 1 pour la résidence urbaine et 2 pour la résidence rurale), ces codes peuvent être enregistrés lors de la saisie ou établis au cours des opérations de vérification, selon les critères définis par l'équipe de vérification. Lorsque l'équipe de vérification fournit une liste des unités géographiques urbaines et rurales, les personnes chargées de traiter les données peuvent facilement attribuer les codes appropriés aux enregistrements concernant l'habitation.

244. Il convient de s'employer à vérifier la cohérence générale des caractéristiques démographiques avec la nature du secteur de dénombrement. Par exemple, dans certains pays, à l'exception des médecins, des enseignants et des personnes exerçant des professions analogues, les zones rurales comptent peu de personnes ayant reçu une formation supérieure et les zones urbaines ne comptent guère de travailleurs agricoles. L'équipe de vérification devrait s'assurer de la classification correcte de la zone géographique.

## B. Contrôles de couverture

### 1. Dénombrement de fait et de droit

245. *Définition d'une résidence habituelle.* D'une façon générale, aux fins d'un recensement, la « résidence habituelle » se définit comme l'endroit où habite une personne au moment du recensement, où elle demeure depuis un certain temps et où elle a l'intention de résider un certain temps. Généralement, la plupart des gens dénombrés n'ont pas déménagé depuis quelque temps, et l'on peut donc définir clairement leur lieu de résidence habituelle. Pour d'autres, l'application de cette définition peut se prêter à de multiples interprétations, surtout si la personne déménage souvent. Il est recommandé que les pays appliquent un seuil de 12 mois avant de considérer une résidence comme habituelle, de sorte que la résidence habituelle serait donc : *a*) le lieu où la personne a vécu continuellement pendant la majeure partie des 12 derniers mois (c'est-à-dire pendant au moins six mois et un jour), sans compter les absences temporaires pour raison de vacances ou de déplacement professionnel, ou a l'intention de vivre au moins six mois; ou *b*) le lieu où la personne a vécu continuellement pendant au moins les 12 derniers mois, sans compter les absences temporaires pour raison de vacances ou de déplacement professionnel, ou a l'intention de vivre au moins 12 mois (Nations Unies, 2008, par. 1.461 à 463).

246. Sur la base de cette définition, les bureaux nationaux de recensement effectuent généralement des recensements de fait (là où les personnes se trouvent le soir du recensement) ou de droit (là où elles se trouvent habituellement). Le contrôle du lien entre les enregistrements concernant l'habitation, en particulier le décompte des personnes qui occupent des logements, et les enregistrements individuels doivent tenir compte du type de recensement. Parfois, les pays recueillent conjointement les informations de fait et de droit. Une rubrique définie pour chaque personne peut indiquer si celle-ci : 1) y réside en permanence; 2) se trouve temporairement en visite, mais habite généralement ailleurs; ou 3) y habite d'ordinaire mais se trouve temporairement absente. Les tableaux de résultats établis sur la base des données de fait recueillies utilisent uniquement les données de type 1 et 2 en présence de données des trois types; les tableaux de résultats établis sur la base des données de droit recueillies utilisent uniquement les données de type 1 et 3 en présence des trois types de données.

247. Les bureaux nationaux de recensement/statistique qui utilisent ces catégories doivent se montrer très prudents dans leur utilisation à chaque stade — lors de la collecte des données, de leur traitement, de leur diffusion et de leur analyse. En particulier, l'établissement de tableaux des ensembles de données doit tenir compte de la catégorie de population désirée. Les utilisateurs de ces trois catégories doivent bien connaître la population choisie, car l'analyse de l'ensemble des données dans leur totalité entraîne la double inscription de certaines personnes. Si l'on a besoin d'une population de fait, lors du dépouillement, il faut exclure la catégorie 3, les personnes temporairement absentes; si l'on a besoin d'une population de droit, il faut exclure la catégorie 2 du dépouillement. Lors des tabulations initiales, l'établissement de tableaux pour les rapports imprimés et d'autres médias, l'équipe de vérification peut décider d'établir un sous-ensemble de données à traiter à partir de l'ensemble des données dans leur totalité. Pour les dépouillements ultérieurs, la documentation des fichiers devrait indiquer clairement comment traiter les diverses possibilités. Des fichiers multiples pourraient être préférables.

248. Le programme de vérification devrait être conçu pour assurer que, en présence des trois types d'enregistrement, les correspondances sont appropriées. Si les enregistrements de fait comportent peu de réponses, cela peut indiquer qu'il s'agit en fait de résidents absents ou qu'un autre problème de dénombrement doit faire l'objet le cas échéant d'un traitement particulier.

## **2. Hiérarchie des ménages et des unités d'habitation**

249. Le chapitre V étudie les relations entre ménages, unités d'habitation et locaux à usage d'habitation. Chaque organisme national de recensement/statistique utilise ces concepts d'une façon qui lui est propre. Toutefois, avant de procéder aux différents contrôles concernant l'habitation, l'équipe de vérification doit définir des méthodes opératoires assurant le respect de la hiérarchie au cours de la collecte et de la saisie des données.

## **3. Fragments de questionnaires**

250. Avant de procéder à une vérification rubrique par rubrique, le programme informatique doit contrôler dans le cadre de la vérification structurelle les enregistrements valides, les enregistrements manquants et les numéros de ligne en double. Il doit en outre déterminer si les enregistrements vérifiés se rapportent à des personnes qui vivent dans des collectivités. Les opérateurs de saisie peuvent faire une erreur d'introduction de données ou incidemment oublier de supprimer des informations fragmentaires (parties d'enregistrement). Les vérifications préliminaires ont notamment pour rôle d'examiner la présence d'enregistrements fragmentaires dans le fichier afin de les détruire. Le cas le plus courant sera celui d'un enregistrement contenant des codes géographiques mais aucune information concernant la population ou l'habitation.

## **C. Structure des enregistrements sur l'habitation**

251. Le nombre de logements dans un immeuble figure parmi les thèmes susceptibles de faire l'objet d'une collecte de données dans le cas de recensements nationaux sur l'habitat. Dans ce cas, l'unité de dénombrement est un immeuble et l'information recueillie se rapporte au nombre de logements classiques de base qui s'y trouvent (voir Nations Unies, 2007, par. 2.524)

252. Le terme « vérification générale » désigne l'opération consistant à assurer que le nombre d'unités d'habitation faisant partie de l'immeuble correspond au nombre total d'unités figurant dans l'enregistrement. S'il n'y a pas concordance, le nombre d'unités d'habitation inscrit comme caractéristique de l'immeuble doit être corrigé pour correspondre au nombre d'unités d'habitation enregistrées. Si le codage de l'immeuble en question indique qu'il comporte 5 unités d'habitation, alors que le décompte réel des unités d'habitation individuelles de cet immeuble est de 4, l'équipe de vérification doit alors décider de l'ajustement à effectuer : *a*) modifier le premier chiffre, ce qui dans la plupart des cas devrait s'avérer plus indiqué; et *b*) introduire une valeur sur la base d'informations concernant les enregistrements existants (ce qu'il conviendrait d'éviter).

## D. Correspondance entre les enregistrements relatifs à l'habitation et à la population

253. Lorsque le recensement ou l'enquête comprend des enregistrements sur l'habitation et sur la population, une vérification de structure s'impose pour s'assurer que les deux types d'enregistrement concordent.

### 1. Habitations inoccupées et occupées

254. À l'inverse d'une unité d'habitation occupée, une unité d'habitation inoccupée ne comporte aucun enregistrement concernant la population. En présence d'enregistrements relatifs à la population, mais lorsque l'unité d'habitation en question figure sur la liste comme étant inoccupée, il faut alors substituer l'état occupé à l'état inoccupé. Parfois, l'enregistrement est organisé de façon à regrouper l'état d'occupation et les « modalités de jouissance » sous la même rubrique, de telle sorte que cette même information doit également être prise en compte lors du choix. S'il existe une réponse concernant la valeur d'une unité occupée par le propriétaire ou le « loyer payé » pour une unité occupée par des locataires, alors les programmes de vérification utilisent ces renseignements pour déterminer la nature du logement; sinon, l'utilisation d'une matrice d'imputation peut s'avérer nécessaire.

255. Lorsqu'une unité d'habitation théoriquement occupée ne fait l'objet d'aucun enregistrement de type démographique, l'équipe de vérification doit décider s'il faut recenser l'unité comme étant inoccupée ou substituer la présence de personnes provenant d'une autre unité. Si l'unité est inoccupée, la modification par imputation de la variable « état d'occupation » ne pose aucune difficulté. Si l'unité est occupée, par contre, l'équipe de vérification doit décider si et comment attribuer les personnes provenant d'une autre unité dans laquelle habite le même nombre de personnes présentant si possible des caractéristiques similaires. Comme il n'est pas possible de connaître les caractéristiques des personnes omises, cette méthode ne devrait être utilisée, pour autant qu'elle puisse l'être, que si l'équipe de vérification convient qu'elle n'a pas d'autre choix. Trois cas possibles sont présentés ci-après :

#### *a*) Décision de laisser une unité d'habitation inoccupée

256. Dans ce cas, l'équipe de vérification décide que les unités d'habitation inoccupées relevées sur le terrain doivent rester considérées comme inoccupées, de telle sorte qu'aucune valeur n'est imputée. Les vérifications dans le cas des unités d'habitation inoccupées sont décrites au chapitre V.

b) Réexamen des unités d'habitation à plusieurs reprises  
afin de compléter les questionnaires

257. Le bureau national de recensement/statistique peut choisir d'appliquer des procédures demandant aux agents de recensement de continuer à réexaminer les données concernant les unités inoccupées jusqu'à ce qu'ils aient la certitude que ces mêmes unités sont effectivement occupées ou inoccupées et jusqu'à ce qu'ils aient recueilli au moins les caractéristiques de base correspondantes. Dans ce cas, l'équipe de vérification doit mettre au point des contrôles permettant de savoir si l'unité est ou non inoccupée ou présente des caractéristiques suffisantes pour être considérée comme « occupée ». Selon le choix de l'équipe de vérification quant aux indications minimales nécessaires, il est procédé soit à la vérification ordinaire décrite au chapitre IV, soit à l'obtention d'enregistrements « donneurs » pour les personnes « manquantes » tel qu'indiqué ci-dessus.

c) Substitution d'une autre unité de logement  
pour les personnes manquantes

258. La description des procédures de substitution de ménages entiers ou de personnes manquantes figure ailleurs dans le présent chapitre. Ces procédures impliquent l'hypothèse selon laquelle les personnes manquantes ont les mêmes caractéristiques que les personnes substituées. Ce qui, de façon presque certaine, n'est généralement pas le cas, et les procédures proprement dites sont particulièrement délicates. Toutefois, sans utiliser les procédures en question, les effectifs de personnes et les nombres de personnes par caractéristique relevée risquent de diminuer.

## 2. Ménages et unités d'habitation figurant en double

259. Les unités d'habitation peuvent figurer en double pour différentes raisons. Parfois, un opérateur de saisie introduit à deux reprises la même unité d'habitation; parfois, des opérateurs de saisie différents réintroduisent accidentellement les mêmes unités d'habitation ou des secteurs de dénombrement entiers, faute d'avoir mis en place un système d'assurance de qualité au niveau du bureau national de recensement/statistique; enfin, un agent de recensement peut très bien enregistrer de façon inadéquate le code géographique d'une unité d'habitation, créant ainsi une information en double en lui attribuant la même identité géographique que celle d'une autre unité.

260. Aucun enregistrement ne figurera en double si les lots introduits sont contrôlés. Toutefois, il conviendrait de mettre au point un programme de vérification garantissant qu'aucun ménage ne figure en double, suite à la saisie en double par erreur du même ménage ou des mêmes ménages. Les pays devraient s'abstenir de trier leurs données, et tant que les problèmes posés par les enregistrements en double ne sont pas éliminés. Avant de procéder au tri, le personnel est en mesure de corriger manuellement les lots; après le tri, il risque par contre de ne pas pouvoir détecter le problème. Lorsque les données sont triées, la présence de ménages enregistrés en double peut être contrôlée au moyen d'une opération de vérification et les données figurant en double par la suite peuvent être éliminées par imputation.

## 3. Ménages et unités d'habitation manquants

261. De même, à la suite du tri, on peut constater l'existence de ménages manquants. Par exemple, le programme de vérification anticipe l'existence d'une séquence de ménages à l'intérieur du niveau géographique le moins élevé, par exemple, 1, 2, 3 et 4,

mais constate uniquement la séquence 1, 2 et 4. Il faut alors décider soit de renuméroter les unités, soit de trouver une méthode acceptable quelconque de substituer une autre unité à l'unité 3 manquante. On dispose de plusieurs moyens d'ajouter des ménages manquants lorsqu'il est clair qu'ils manquent effectivement et doivent être rajoutés. Selon une méthode, on peut décider simplement de faire figurer en double le ménage précédent. Toutefois, si l'on connaît le nombre de personnes composant ce ménage, comme c'est souvent le cas (même si leurs caractéristiques ne sont pas connues), on peut retourner en arrière et introduire en double la précédente unité ayant le même nombre de membres. De même, si l'on connaît l'âge et le sexe des membres du ménage, cette information peut faciliter le choix d'une unité de substitution. Il n'est pas jugé prudent de tenter d'utiliser l'imputation dynamique pour créer des informations sur les membres du ménage car cette méthode produit souvent des variables incompatibles les unes avec les autres.

#### **4. Correspondance entre le nombre d'occupants et la somme des occupants**

262. Le nombre d'occupants inscrits dans l'enregistrement relatif à l'habitation doit être exactement égal au décompte des personnes du ménage. Le programme de vérification additionne donc le nombre de personnes, puis compare cette valeur au nombre d'occupants inscrits dans l'enregistrement relatif à l'habitation. Si ces deux valeurs diffèrent de la valeur correspondant au nombre d'occupants, il faut alors ajuster soit le nombre d'occupants de façon à ce qu'il soit identique au décompte des personnes, soit les données individuelles. Le chapitre V décrit de façon détaillée le processus de vérification des données d'habitation en ce qui concerne le nombre d'occupants.

##### **a) Quant le nombre d'occupants est supérieur au décompte des occupants**

263. Si la valeur attribuée au nombre d'occupants figurant dans l'enregistrement concernant l'habitation est supérieur au décompte des enregistrements individuels, l'équipe de vérification est alors confrontée à une véritable difficulté. Personne ne peut connaître les caractéristiques des personnes manquantes. Par conséquent, les équipes de vérification qui choisissent d'imputer les personnes manquantes caractéristique par caractéristique ou en les remplaçant par des personnes qui vivent dans des ménages similaires risquent de se trouver face à un dilemme. Les personnes manquantes ne doivent pas être remplacées. En revanche, si le nombre d'occupants est accepté, la solution de remplacement consiste à diminuer l'effectif de la population dénombrée. L'équipe de vérification doit étudier la situation dans son ensemble puis décider de la voie à suivre.

264. Il existe plusieurs moyens pour repérer et remplacer des enregistrements manquants, mais aucun d'eux n'est parfaitement satisfaisant. Il est possible de sauvegarder des enregistrements de ménages complets comportant des variations des principales caractéristiques. Lorsqu'on trouve un ménage comportant un certain nombre d'individus, mais pas la totalité, il est possible de rechercher dans le jeu de données un ménage dont la plus grande partie sinon la totalité des caractéristiques connues coïncident, ce qui permet ensuite d'ajuster les caractéristiques des personnes manquantes d'après celles des autres personnes du ménage donneur. Cependant, la programmation de cette opération est extrêmement complexe, et les bureaux nationaux de recensement/statistique qui utilisent cette méthode doivent préparer longtemps à l'avance sa mise en œuvre.

265. Une variante de cette procédure consiste à associer des indicateurs à tous les ménages comportant des enregistrements manquants puis à effectuer le reste des vérifications. Au terme du processus de vérification, après avoir corrigé toutes les données individuelles, l'équipe de vérification peut choisir de demander aux informaticiens d'effectuer sur le fichier les ajouts et les modifications possibles au moyen du jeu de données complet dûment vérifié; selon cette approche descendante, l'équipe est en mesure d'identifier les donneurs acceptables.

**b) Contrôle du nombre de personnes par sexe**

266. Parfois, l'enregistrement concernant l'habitation porte la mention du nombre d'occupants par sexe. Le programme de vérification doit alors additionner séparément le nombre de personnes de chaque sexe. Là encore, si la somme diffère des nombres respectivement indiqués, l'une des deux valeurs doit être ajustée dans chaque cas. D'ordinaire, on préfère ajuster les totaux inscrits dans les enregistrements relatifs au logement, plutôt que d'ajouter des enregistrements manquants ou de supprimer des enregistrements comportant des informations utiles parce que l'agent recenseur risque d'avoir commis une erreur sur le formulaire relatif au logement.

**c) Numérotation séquentielle**

267. Les enregistrements relatifs à la population devraient être numérotés dans un ordre séquentiel. Ces numéros devraient figurer sur le questionnaire comme une variable (telle qu'un numéro de ligne ou un numéro séquentiel). En outre, les numéros séquentiels devraient apparaître dans l'ordre numérique. Des erreurs peuvent néanmoins se produire; parfois, les questionnaires sont déclassés parce que les agents de recensement regroupent les informations dans le mauvais ordre, ou sautent des pages, laissant involontairement des pages vides dans le jeu de données. Bien qu'une lacune de numérotation séquentielle n'affecte ni les opérations de vérification ni celles de mise en tableaux, nombre de bureaux nationaux de recensement/statistique choisissent de reclasser les individus dans l'ordre adéquat. Le programme de vérification doit donc pouvoir identifier les personnes qui ne sont pas dans le bon ordre et les renuméroter. Puisque la renumérotation affecte parfois le lien avec le chef de ménage, les spécifications de vérification doivent être réexaminées. Ainsi, la renumérotation affectera certainement des variables telles que le numéro de ligne de la mère ou du mari.

## **5. Correspondance entre les occupants et le type d'immeuble/de ménage**

268. Le type de lien entre les membres d'un ménage doit être compatible avec le type d'unité d'habitation. Parfois, les membres d'un ménage figurent dans une habitation déclarée en tant que logement collectif ou vice versa. En l'occurrence, le type de lien ou le type d'unité d'habitation doit nécessairement tenir compte de la taille du ménage et de différentes variables.

## **E. Enregistrements en double**

269. Les questionnaires à la lecture optique ou autre ne sont guère susceptibles de présenter des numéros de ligne en double. Pour les formulaires dont le contenu doit être introduit au clavier, le bureau national de recensement/statistique peut convenir de vérifier la correspondance entre la liste des ménages et les numéros de ligne du

ménage à introduire manuellement. Ce contrôle manuel peut améliorer la qualité des données saisies, en particulier si l'on rapproche : 1) les noms des personnes figurant sur une page comportant la liste de toutes les personnes du ménage; et 2) les données figurant dans les colonnes, rangées ou pages relatives aux personnes. Ainsi, deux personnes qui à première vue semblent constituer le dédoublement d'une seule peuvent en fait être des jumeaux si l'on se réfère à leurs noms.

270. Les formulaires remplis au clavier ne doivent pas comporter de numéros de ligne en double si les écrans de données et les instructions de saut de champ sont correctement configurés. La plupart des progiciels actuels créent automatiquement des numéros séquentiels dans le cadre du processus de saisie. Or, une erreur peut être introduite en cas de saisie d'enregistrement en double pour une personne ou encore de création d'un enregistrement double dû à un numéro de ligne erroné. Au fur et à mesure du traitement de chaque enregistrement, le programme de vérification le compare aux précédentes données de population relatives à l'unité d'habitation. La vérification effectuée doit contrôler la saisie correcte de chaque numéro de ligne. Les numéros de ligne en double constituent des erreurs et doivent être modifiés.

271. Au lieu d'utiliser un progiciel du commerce, les pays peuvent choisir de mettre au point leurs propres systèmes d'introduction des données. Dans ce cas, l'équipe de vérification doit convenir du niveau acceptable d'erreurs. Parmi les approches envisageables à cet effet, l'une d'elles pourrait consister à suivre les principes ci-dessous :

- a) Si deux enregistrements différents correspondent au même numéro de ligne et si le nombre de caractéristiques différentes est inférieur ou égal à deux, le programme de vérification doit éliminer l'un des enregistrements puisqu'il s'agit vraisemblablement d'un double;
- b) Si le nombre de caractéristiques différentes est supérieur ou égal à trois, le numéro de ligne doit être modifié<sup>7</sup>.

7 Par le passé, les enregistrements en double étaient traditionnellement détectés et corrigés à la main, mais de plus en plus ces opérations sont au moins en partie automatisées. Une étude récente (Winkler, 2006) commence à envisager d'automatiser à la fois les vérifications de structure et de contenu.

## F. Populations spéciales

### 1. Personnes vivant en collectivité

272. La vérification structurelle doit traiter les personnes vivant dans des collectivités telles qu'institutions, casernes ou maisons de soins différemment de celles qui vivent dans des unités d'habitation normales. Puisque dans les collectivités il n'existe généralement pas de chef de ménage, les pays doivent décider de la meilleure façon de distinguer les différents types d'unités. Une méthode consiste à définir un type d'enregistrement propre aux collectivités; une autre, à attribuer au lien avec le chef de ménage un code particulier correspondant aux logements en groupe ou collectifs.

- a) Cas dans lesquels les enregistrements collectifs constituent un type d'enregistrement différent

273. Si le bureau national de recensement/statistique décide d'utiliser un type d'enregistrement distinct, l'équipe de vérification n'aura aucune difficulté à déterminer quels enregistrements sont de type collectif. Les tableaux correspondants sont aisément réalisables par simple référence directe aux enregistrements collectifs, (par exemple, le type de collectivité) et peuvent faire l'objet d'opérations distinctes de vérification et d'imputation. Quant aux variables exclues des enregistrements collectifs, il est facile de les vérifier en s'assurant que les données en question sont effectivement

laissées en blanc. Il en résulte toutefois un fichier plus encombrant, car ces enregistrements sont généralement plus courts que les enregistrements normaux, mais ils occupent autant de place que dans un fichier de type rectangulaire. De plus, au cours de la vérification et de l'imputation, certains programmes peuvent avoir à vérifier certaines rubriques tant des enregistrements de population que des enregistrements collectifs.

**b) Cas dans lesquels une variable distingue les enregistrements collectifs des autres enregistrements**

274. Lorsque l'on utilise une variable distincte de préférence à un type d'enregistrement distinct, l'équipe de vérification peut avoir plus de difficultés à identifier les enregistrements de type collectif. Dans ce genre de situation, les tableaux relatifs aux collectivités sont également faciles à établir en se référant à la variable proprement dite, qui indique quels enregistrements se rapportent à des personnes faisant partie de collectivités. Les variables propres aux collectivités (comme le type de collectivité) peuvent également faire l'objet de vérifications et d'imputations distinctes. Il est facile de vérifier les variables exclues des enregistrements collectifs pour s'assurer qu'elles sont effectivement laissées en blanc en se référant au code des enregistrements collectifs. Il en résulte un fichier plus compact, puisque les enregistrements supplémentaires concernant les personnes des collectivités ne sont pas nécessaires, mais figurent simplement en tant qu'enregistrements de données démographiques, dans lesquels la variable ménage/collectivité porte un code différent. Lors des opérations de vérification et d'imputation, le programme devra contrôler uniquement certaines rubriques des enregistrements de population et non des deux types d'enregistrement (données de population et données sur les collectivités).

**c) Cas dans lesquels le code « type de collectivité » est omis**

275. Le code indiquant le type de collectivité est parfois manquant ou non valide, ou encore il peut y avoir un défaut de concordance entre le code de collectivité et les codes de lien avec le chef de ménage. En l'absence du code de type de collectivité, mais si les codes de lien avec le chef de ménage indiquent qu'il s'agit bien d'une collectivité, la solution proposée consiste à modifier en conséquence le codage du type de collectivité. Si ce code est présent mais si le code du lien avec le chef de ménage est absent, ce dernier peut être établi en fonction du type de collectivité.

**d) Cas dans lesquels le code « type de collectivité » est présent, mais où il existe un lien entre toutes les personnes**

276. Si l'enregistrement comporte un code correspondant à une collectivité, mais si les codes « lien avec le chef de ménage » indiquent l'existence d'un lien entre toutes les personnes de l'unité d'habitation, il convient alors de modifier le type de collectivité pour indiquer qu'il s'agit d'une unité d'habitation. En revanche, si l'unité d'habitation est codée en tant que ménage mais qu'il n'y a pas deux personnes vivant dans cette unité entre lesquelles il existe un lien, il est parfois nécessaire d'attribuer à cette unité le code d'habitat de groupe ou collectif. Par ailleurs, un ménage peut compter 5 ou 6 personnes entre lesquelles il n'existe pas de lien sans toutefois constituer un habitat collectif. Tel qu'indiqué plus haut, les membres de l'équipe de vérification devront parfois se consulter pour résoudre des cas particuliers inhabituels.



e) Distinction entre différents types de collectivité

277. La plupart des pays distinguent différents types de collectivité. Ils ventilent souvent l'information en différents types particuliers de logement collectif. Ces informations peuvent soit être codées séparément en tant que type de logement collectif, soit figurer en tant que choix de valeurs des codes de lien avec le chef de ménage.

## 2. Groupes difficiles à dénombrer

a) Migrants saisonniers

278. Dans certains pays où l'on observe des migrations saisonnières, l'interviewer devra savoir si l'unité est inoccupée ou occupée en fonction de la référence temporelle. Dès lors, même si les données concernant le ménage sont complètes, ce dernier peut également être compté (dénombré) dans un autre lieu. Évidemment, le contraire peut également se produire : un ménage possédant deux logements situés en des lieux distincts (ces personnes vivent de préférence en des lieux différents selon la saison) risque également d'être omis si l'on n'y prend garde.

279. Parfois, des ménages entiers vivent très régulièrement une partie de l'année à un endroit et le reste du temps ailleurs. Le bureau national de recensement/statistique ainsi que l'équipe de vérification doivent décider du mode de traitement de divers types de situations. Par exemple, certaines personnes passent régulièrement une partie de l'année dans une autre maison : Les gens qui vivent dans la partie la plus froide du pays pendant la saison chaude et dans la partie au climat plus clément pendant la saison froide. Un autre cas est celui des itinérants qui voyagent une partie de l'année mais sont sédentaires le reste du temps, ce qui peut coïncider avec l'époque à laquelle le pays décide de procéder au recensement.

b) Personnes sans abri

280. Par définition, l'enregistrement d'une personne sans abri ne contiendra aucune information concernant l'habitation. Toutefois, la création d'un enregistrement « fictif » (nouvel enregistrement qui comprend initialement des valeurs laissées en blanc pour certaines variables) facilitera les vérifications structurelles et permettra d'obtenir un enregistrement cohérent avec la structure des autres unités d'habitation. L'équipe de vérification devra décider éventuellement de créer cet enregistrement fictif d'habitation pour faciliter le traitement et la mise en tableaux des données.

c) Itinérants et personnes vivant dans des régions difficiles d'accès

281. Là encore, de même que pour les personnes sans abri, il peut être très difficile d'effectuer une vérification structurelle. Certains pays collectent certaines informations sur « le logement », de sorte que ces informations puissent être utilisées pour aider à vérifier la structure de « l'unité ». De cette façon, les vérifications concernant le logement seraient différentes de celles utilisées pour les unités ordinaires. Les informations démographiques devraient être recueillies comme elles le seraient pour les personnes vivant dans des logements normaux et vérifiées de la façon habituelle, selon les directives ci-dessous.

d) Résidents civils temporairement absents de leur pays

282. Dans les recensements de droit, les résidents civils temporairement absents de leur pays mais vivant au sein de ménages qui peuvent les enregistrer devraient

être inclus dans les vérifications normales concernant la population. Pour les recensements de droit, un certain indicateur devrait identifier les personnes qui sont temporairement absentes afin de permettre de déterminer à la fois la population de droit et la population de fait. La vérification concernant le logement ne présentera guère de différences du fait des absents. Bien entendu, toutefois, ces personnes ne seront pas incluses dans un recensement de fait; par conséquent, elles ne figureront pas dans les vérifications concernant la population.

e) **Civils étrangers qui ne franchissent pas quotidiennement la frontière et se trouvent dans le pays à titre temporaire, y compris les personnes sans papiers ou les passagers de navires se trouvant dans un port au moment du recensement**

283. Étant donné que, pour un recensement de fait, quiconque réside dans le pays au moment du recensement devrait être recensé, ces personnes devraient également être incluses dans le recensement. Ces personnes devraient être incluses à l'endroit où elles résident au moment du recensement, et leur vérification devrait se faire selon les vérifications effectuées normalement pour les données concernant la population. Si le logement n'est pas recensé comme logement collectif ou sous une autre forme non habituelle d'unité d'habitation, la vérification n'aura pas lieu non plus pour ces personnes. Si les navires se trouvant dans un port sont considérés comme unités d'habitation, les caractéristiques de ces unités devraient être décrites sur la base d'informations recueillies pour d'autres navires pour les imputations dynamiques.

284. Les étrangers qui se trouvent dans le pays uniquement à titre temporaire ne sont pas censés être inclus dans les recensements de droit. Les personnes sans papiers y seraient incluses, en particulier dans les pays qui ne recensent pas séparément les personnes avec et sans papiers (encore que le fait de faire une telle distinction produirait normalement de meilleurs résultats). Les personnes en transit ne seraient pas incluses dans le recensement de droit, à moins qu'elles ne soient en transit dans la localité mais résident habituellement dans le pays. Si un navire touche régulièrement un port du pays, les personnes se trouvant à bord de ce navire devraient probablement être recensées comme résidents habituels et, à ce titre, faire l'objet d'une vérification.

f) **Réfugiés**

285. Des réfugiés peuvent occuper des logements temporaires et nécessiter alors une indication de cette situation par l'affectation d'une valeur à une variable particulière, un type d'enregistrement distinct ou un enregistrement fictif de leur logement pour rendre compte de cette situation. Normalement, les rubriques logement et population utilisent la vérification courante, et les imputations dynamiques font état de « logement pour réfugiés » à titre d'indicateur.

g) **Personnel militaire, naval et diplomatique et membres de leur famille se trouvant hors du pays et personnel militaire, naval et diplomatique étranger et membres de leurs famille se trouvant dans le pays**

286. Pour un recensement de droit, le personnel militaire, naval et diplomatique et les membres de leurs familles se trouvant aussi bien à l'intérieur qu'à l'extérieur du pays seraient normalement inclus dans le recensement. Pour nombre de pays, les informations sur le personnel militaire ne figurent pas dans le recensement et le bureau national de statistique doit procéder à des comptages simples ou à des compta-

ges effectués à partir d'un minimum d'autres informations. Lorsque les informations sont limitées, l'utilisation d'imputations dynamiques s'avère difficile, et des erreurs risquent d'être introduites dans le jeu de données; il est donc généralement préférable de ne pas inclure les familles de militaires dont il est rendu compte de cette façon dans le recensement. Le personnel diplomatique se heurte parfois à des problèmes similaires. Toutefois, le dénombrement effectué à l'intérieur d'un pays peut produire de bons résultats lorsque l'on utilise des questionnaires et des procédures types; par conséquent, ces unités d'habitation devraient être incluses dans la vérification simple, mais un indicateur devrait signaler la situation spéciale de ces unités. Vu que les unités d'habitation situées hors du pays risquent de ne pas être dénombrées de la façon habituelle, faudra-t-il prendre un soin particulier lorsqu'il s'agira de décider d'inclure ou non ces unités dans les vérifications ? (Elles pourraient quand même être incorporées à certains tableaux).

287. Pour un recensement de fait, on n'inclut généralement que les unités d'habitation se trouvant dans le pays. Normalement, les familles de personnel militaire, naval et diplomatique vivant hors du pays ne seraient pas recensées.

*h) Civils étrangers qui franchissent quotidiennement la frontière pour travailler dans le pays*

288. Normalement, les civils étrangers qui franchissent quotidiennement la frontière pour venir travailler dans le pays ne seraient inclus ni dans un recensement de droit ni dans un recensement de fait, parce qu'ils ne résident pas dans le pays à la date de référence, pas plus qu'ils n'y résident habituellement. Ils seraient normalement inclus dans les recensements de droit et de fait effectués dans le pays d'où ils viennent.

*i) Résidents civils qui franchissent quotidiennement la frontière pour travailler dans un autre pays*

289. Les résidents civils qui franchissent quotidiennement la frontière pour aller travailler dans un autre pays sont résidents du pays qui effectue le recensement et devraient donc figurer à la fois dans les comptes de droit et de fait. Les données les concernant et concernant leur logement feraient l'objet d'une vérification selon la procédure normale.

*j) Membres d'équipage de navires marchands et pêcheurs résident du pays se trouvant en mer au moment du recensement (y compris ceux qui n'ont pas d'autre lieu de résidence que leurs quartiers à bord d'un navire)*

290. Les personnes appartenant à la marine marchande seraient recensées dans un pur recensement de droit, de même que dans un recensement de droit modifié (ajusté pour inclure les personnes n'ayant pas d'autre lieu de résidence), mais pas dans un recensement de fait. Lorsqu'elles sont incluses, les vérifications concernant le logement doivent faire état du type spécial de résidence, mais il devrait être possible d'utiliser des vérifications types pour les données démographiques lorsque l'on utilise le questionnaire normal du pays à bord des navires.

## G. Détermination du chef de ménage et du conjoint

### 1. Détermination de la variable chef de ménage

291. Pour identifier les membres d'un ménage, il est d'usage d'identifier d'abord le chef de ménage ou le membre de référence, puis les autres membres, en fonction de leur lien avec lui. Le chef de ménage est la personne reconnue comme telle par les autres membres. Les pays peuvent utiliser le terme qu'ils jugent le plus approprié pour l'identifier (chef de ménage, membre de référence du ménage, etc.), cette personne servant uniquement à déterminer les liens entre les membres du ménage. Il est souhaitable que chaque pays présente dans les rapports qu'il publie les notions et les définitions utilisées (Nations Unies, 2007, par. 2.114).

#### a) Ordre des liens avec le chef de ménage

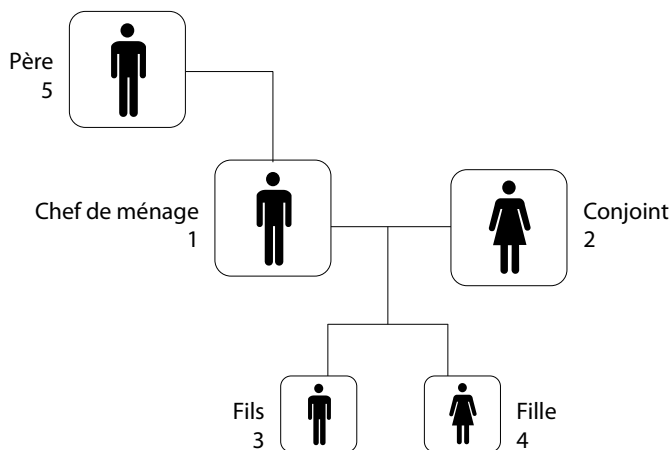
292. L'ordre des liens avec le chef de ménage au sein de l'unité d'habitation a une incidence sur les vérifications; en effet, dans nombre de cas, celles-ci supposent que le chef de ménage est la première personne et vérifient en premier lieu les données le concernant. Par exemple, des variables telles que la langue, le groupe ethnique et la religion sont contrôlées en premier lors de la vérification concernant le chef de ménage. Si pour l'une ou l'autre de ces variables, les informations sont valides, celles-ci sont imputées à toute autre personne du ménage en cas d'omission, d'erreur de codage ou d'erreur d'introduction au clavier (voir chapitre IV). Il faut donc vérifier en priorité les données du chef de ménage, ses caractéristiques servant ensuite à attribuer ou à imputer des valeurs aux autres membres du ménage.

#### b) Cas dans lesquels le chef de ménage n'est pas la première personne

293. Les décisions prises sur le terrain par les agents de recensement en fonction des différentes situations rencontrées quant à l'identification du chef de ménage affectent le processus de vérification. Pour mieux comprendre le problème, examinons tout d'abord le cas du ménage présenté à la figure 25.

Figure 25

Exemple de ménage dont le chef figure en premier sur la liste

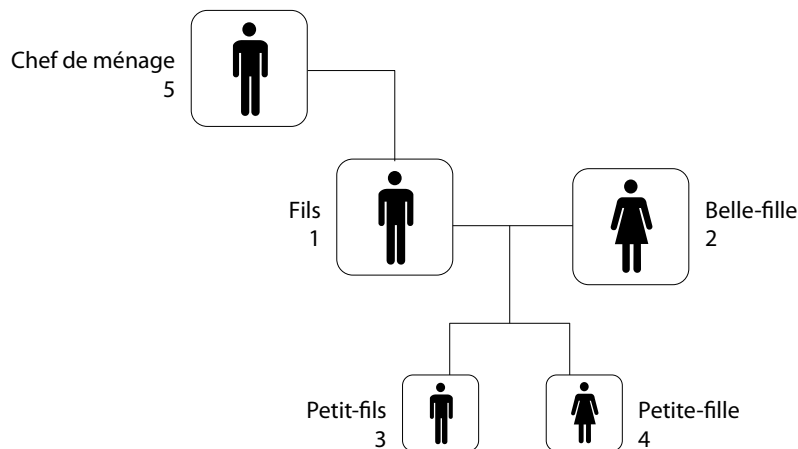


294. Ce ménage correspond aux cas types rencontrés sur le terrain : un chef de ménage et un conjoint, leurs enfants, ainsi que le père du chef de ménage. Si l'agent de recensement recueille des informations de cette façon, une vérification fondée sur l'inscription du chef de ménage en tête de liste du ménage ne posera pas de problème.

295. Toutefois, si le recensement se déroule de telle sorte que le grand-père est désigné comme étant le chef de ménage, les liens au sein du ménage sont modifiés, selon le schéma de la figure 26. Tel serait le cas si un agent de recensement se présentait dans une maison, y trouvait une famille nucléaire constituée d'un mari, d'une femme et de deux enfants et si, pendant l'interview, le père du chef de ménage était dans la pièce et déclarait être le chef de ménage. Moyennant l'accord du chef de ménage présumé, la personne n° 5 deviendrait alors le chef de ménage et la personne 1, le fils, la personne 2, la belle-fille, etc.

Figure 26

**Exemple de ménage dont le chef est classé en cinquième position**



296. Comme le montre évidemment le cas de ces deux ménages, la désignation de tel ou tel chef de ménage conduit à des chemins de vérification différents. La détermination du chef de ménage effectif pour le reste des opérations de vérification et de mise en tableaux donne lieu à trois possibilités distinctes : *a*) un pointeur (indicateur) peut signaler la personne considérée comme le chef de ménage, ce pointeur pouvant servir tout au long des opérations de vérification et de mise en tableaux; *b*) si le chef de ménage ne figure pas en première position sur la liste, il est possible de le reclasser en première position et de modifier en conséquence la position des différentes personnes qui le précédaient; ou *c*) les codes attribués à la variable lien avec le chef de ménage peuvent être changés de façon à ce que la personne placée en premier soit le chef de ménage, indépendamment des autres liens.

*i) Attribution d'un pointeur à l'enregistrement du chef de ménage*

297. Dans le cadre des procédures de vérification concernant le chef de ménage, le numéro de ligne du chef de ménage au sein de l'unité d'habitation est déterminé au moyen d'un pointeur. Si le chef de ménage conserve le numéro de classement relevé, un pointeur peut être associé à ce numéro, pour pouvoir identifier facilement le chef de ménage chaque fois qu'une vérification ou une mise en tableaux particulière l'exigera. Une variable « pointeur du chef de ménage » peut être mise à une valeur égale au nu-

méro de ligne du chef de ménage et utilisée au cours de la vérification pour procéder à l'imputation des caractéristiques manquantes ou non valides concernant les autres personnes de l'unité d'habitation. Si le chef de ménage est la personne portant le n° 1, la valeur de la variable « pointeur du chef de ménage » est égale à 1.

ii) *Désignation du chef de ménage en tant que personne n° 1*

298. L'équipe de vérification peut décider de classer le chef de ménage en tant que personne n° 1. Toutefois, la programmation de cette opération s'avère légèrement plus complexe par comparaison à l'opération requise dans le cas i). Les informaticiens doivent mettre au point un programme qui amène le chef de ménage en première position sur la liste des membres, où il précède désormais la personne qui s'y trouvait au préalable, puis la personne qui était en deuxième position, jusqu'à la personne qui précédait immédiatement le chef de ménage. Donc, le classement des différents membres passe de 1, 2, 3, 4, 5 à 5, 1, 2, 3, 4. Une fois la modification faite, le chef de ménage occupe la position n° 1, ce qui facilite les vérifications suivantes, puisque le chef de ménage conservera ce numéro. Cette opération entraîne toutefois une certaine « détérioration » de l'intégrité du jeu de données. En effet, l'ordre de classement des personnes ayant été modifié, les analystes risquent d'avoir des difficultés pour établir l'ordre réel dans lequel les réponses ont été recueillies sur le terrain et l'incidence potentielle de cet ordre sur l'interprétation des résultats.

iii) *Réattribution de codes de lien avec le chef de ménage de façon à placer celui-ci en première position*

299. Si l'équipe de vérification décide que la première personne de la liste doit être le chef de ménage, il faut alors intégrer les tâches *a* et *b* aux procédures de contrôle :

- a) Le code affecté au chef de ménage est attribué à la première personne;
- b) Un programme réattribue les valeurs de codage aux autres personnes du ménage afin de modifier le classement.

300. Par exemple, à la figure 26, le parent commence en tant que chef de ménage. Si la qualité de chef de ménage est attribuée à la personne n° 1, il faudra attribuer la qualité de conjoint à la personne n° 2, puis celle d'« enfant » aux personnes n° 3 et 4 et réattribuer la qualité de « parent » à la personne n° 5 (tel qu'indiqué à la figure 25). Le sous-programme contiendra alors une matrice regroupant les valeurs initiales et les valeurs modifiées.

301. Cette procédure affecte encore davantage l'intégrité du jeu de données. Le classement des personnes n'est pas modifié comme il l'était dans l'exemple précédent, et les analystes n'auront aucun mal à déterminer l'ordre des personnes dans lequel les informations ont été effectivement recueillies sur le terrain. Les liens avec le chef de ménage seront alors tous modifiés et les analystes ignoreront qui était choisi initialement comme chef de ménage. En outre, si le numéro d'ordre de la mère, celui du père ou celui du conjoint est recueilli lors du recensement ou de l'enquête, il faut en tenir compte dans toute opération de renumérotation. En revanche, l'exploitation des données sera formellement facilitée avec le chef de ménage en première position. Contrairement à l'exemple précédent, cette procédure n'exige pas des programmeurs qu'ils permutent les enregistrements.

### c) Plusieurs chefs de ménage

302. En présence de plusieurs chefs de ménage, il incombe à l'équipe de vérification de déterminer qui doit être considéré comme tel. La vérification doit s'appuyer sur les caractéristiques définies par les spécialistes du sujet et d'autre part sur les schémas de vérification adoptés. Le programme de vérification doit ensuite réattribuer la valeur de la variable lien avec le chef de ménage de la personne ou des personnes précédemment identifiée(s) en tant que chef de ménage.

303. Certains pays acceptent comme cas particulier l'existence de « cochefs de ménage » soit en raison de conditions socioéconomiques particulières (notamment d'absences fréquentes de l'homme travaillant dans le secteur des industries extractives ou à d'autres activités qui le tiennent éloigné et l'obligent à déléguer la fonction de chef de ménage à son conjoint) ou parce que les personnes interrogées tiennent à « l'égalité ». Traditionnellement, pour les besoins de la vérification, il importe qu'une personne et une seulement soit désignée comme chef de ménage et à ce que, en pareil cas, les données initiales soient conservées dans l'enregistrement. Néanmoins, les Nations Unies (2008, par. 2.117) prévoient une disposition pour partage de la fonction de chef de ménage. Si un pays choisit d'accepter la présence de cochefs de ménage, ceux-ci doivent figurer comme tels dans la vérification; toutefois, nombre des vérifications ultérieures suggérées dans le présent manuel auraient également été modifiées : lorsque les cochefs de ménage pratiquent différentes religions, appartiennent à des tribus différentes ou présentent différentes autres caractéristiques démographiques et sociales, une seule personne ne peut plus être utilisée lors des procédures d'imputation.

### d) Absence de chef de ménage

304. De manière analogue, en l'absence de chef de ménage, il incombe à l'équipe de vérification de déterminer qui nommer en tant que tel. Dans ce cas, les relations entre les autres personnes du ménage devront vraisemblablement être modifiées au cours de la vérification. Pour déterminer qui doit être nommé chef de ménage, il convient de tenir compte de variables telles que l'âge, le niveau d'instruction et l'activité économique. L'annexe IV présente un organigramme d'un exemple de vérification pour l'identification du chef de ménage.

## 2. Vérification concernant le conjoint

### a) Cas d'un conjoint unique dans des sociétés monogames

305. Si l'enregistrement comporte un conjoint et un seul, la variable « pointeur de conjoint » enregistre le numéro de ligne du conjoint en vue des vérifications ultérieures. Ces vérifications pourraient consister notamment à s'assurer que le chef de ménage et le conjoint sont de sexes opposés, qu'ils présentent une différence d'âge appropriée ou d'autres caractéristiques pertinentes. (Dans les pays où se pratiquent les mariages homosexuels, il conviendra d'ajuster la vérification en conséquence).

### b) Cas de plusieurs conjoints dans les sociétés monogames

306. Dans les sociétés monogames, lorsque le jeu de données indique l'existence de plusieurs conjoints, une vérification est nécessaire pour déterminer qui est le conjoint et redéfinir les valeurs « lien avec le chef de ménage » des autres personnes identifiées en tant que conjoints. Là encore, les spécialistes du sujet doivent déterminer les caractéristiques et le schéma de traitement des vérifications.

### c) Conjoints dans les sociétés polygames

307. Dans une société polygame, si le jeu de données fait apparaître plusieurs conjoints, l'équipe de vérification peut décider de laisser l'information inchangée ou d'effectuer différents contrôles de cohérence. Il conviendrait au minimum de vérifier si le chef de ménage et chacun des conjoints sont de sexe opposé. Si cette opération révèle l'existence de conjoints du même sexe que le chef de ménage, il convient alors d'effectuer la vérification précédente pour ces conjoints.

### d) Autres caractéristiques des chefs de ménage et des conjoints

308. Les bonnes pratiques de vérification comportent l'imputation d'autres données importantes pour le chef de ménage et le conjoint lorsque ces données font partie de la vérification générale. Parmi ces données figurent l'âge du chef de ménage et du conjoint et leur état matrimonial, qui peuvent s'avérer nécessaires par la suite pour les fichiers d'imputation et à d'autres fins de vérification. En outre, il est également utile de recueillir dès le début des données « sociales » telles que la religion, l'appartenance ethnique et la langue du chef de ménage, surtout si le chef de ménage n'est pas indiqué comme personne n° 1. Comme la plupart des programmes commencent par la première personne, puis passent aux suivantes, il est important de disposer des informations concernant le chef de ménage avant de vérifier celles concernant les autres personnes vivant dans l'unité d'habitation. Un exemple d'organigramme de vérification concernant le conjoint est présenté à l'annexe IV.

## H. Âge et date de naissance

### 1. Indication de la date de naissance, mais omission de l'âge

309. Lorsque la date de naissance est indiquée mais non l'âge, il est possible d'obtenir cette dernière information en soustrayant la date de naissance de la date du recensement ou de l'enquête. Certains bureaux nationaux de recensement/statistique conviennent de définir l'âge en fonction uniquement de l'année du recensement et de l'année de la naissance, ce qui donne une valeur pouvant s'écarter de la valeur réelle. En cas d'utilisation de l'année et du mois, l'âge sera plus précis, mais la précision optimale s'obtient quand on utilise le jour, le mois et l'année de naissance.

### 2. Défaut de concordance de l'âge et de la date de naissance

310. Lorsque l'âge et la date de naissance sont relevés dans le cadre du recensement ou de l'enquête, on obtient l'âge en soustrayant la date de naissance de la date de référence. Si cet âge diffère de plus d'un an de l'âge déclaré, l'équipe de vérification peut décider d'introduire une correction. Normalement, la date de naissance prévaut sur l'âge déclaré et l'âge calculé est alors substitué à l'âge déclaré.

## I. Décompte des données invalides

311. Certaines équipes de vérification décident parfois de mettre en œuvre des procédures de comptage du nombre de données non valides ou incohérentes concernant les principales variables (ou la totalité des variables), telles que l'âge et le sexe, avant de passer aux opérations de vérification proprement dites. L'équipe dispose parfois de plusieurs matrices d'imputation dynamique lorsqu'elle a effectué un travail de



préparation préalable, ou lorsqu'elle réalise des enquêtes périodiques portant sur les mêmes rubriques. Lorsque le pourcentage de données non valides ou incohérentes est très faible, elle peut décider d'utiliser seulement un petit nombre de variables pour l'imputation. Par contre, si le pourcentage d'erreurs est plus élevé, l'utilisation d'un plus grand nombre de variables peut s'avérer nécessaire, compte tenu du plus grand nombre d'imputation requises.

312. Les matrices d'imputation de taille plus réduite sont généralement préférables puisqu'elles sont plus simples à contrôler au fur et à mesure de la mise au point des vérifications et des imputations et sont également d'un usage plus facile au cours de la vérification proprement dite. Si les valeurs sont employées plusieurs fois, une matrice d'imputation plus grande et mieux adaptée s'avérera néanmoins indispensable.



## Chapitre IV

# Vérifications concernant les données sur la population

313. Le chapitre IV traite des vérifications des données sur la population, notamment des données liées aux caractéristiques démographiques, migratoires, sociales et économiques. Les spécifications correspondantes prennent en considération la validité de chaque donnée et la cohérence, d'une part, entre les données sur la population et, d'autre part, entre les données sur la population et sur l'habitation. Les connaissances disponibles quant aux liens entre les éléments d'information permettent de concevoir des vérifications de cohérence propres à garantir une meilleure qualité des données à mettre en tableaux. Par exemple, il ne devrait pas y avoir, dans les enregistrements de données sur la population, de femmes âgées de 15 ans et mères de 10 enfants ou d'enfants de 7 ans inscrits à l'université.

314. Lors de l'attribution de valeurs aux données manquantes ou non valides concernant la population, l'équipe de vérification doit décider soit de leur attribuer la valeur « non indiqué », soit de procéder par imputation statique (*cold deck*) des valeurs inconnues ou autres, ou encore par imputation dynamique (*hot deck*), d'après les caractéristiques d'autres personnes ou d'autres unités d'habitation.

315. Dans bien des cas, la solution de l'imputation dynamique est choisie de préférence car elle supprime les vérifications au stade de l'exploitation, lorsque les décisions à prendre concernant les valeurs inconnues ne peuvent s'appuyer que sur le contenu des tableaux proprement dits. Les matrices d'imputation fournissent des données permettant de compléter les omissions, de remplacer les données non valides ou de résoudre les incohérences lorsqu'on ne dispose d'aucune rubrique connexe comportant des réponses valides. Dans certains cas, les caractéristiques démographiques présentent une diversité notable au niveau national mais beaucoup plus réduite au niveau de la plupart des localités. Dans d'autres pays, les variations peuvent être considérables d'une localité à l'autre, en particulier en matière de résidence urbaine et rurale. Il faut prendre en compte cette variabilité lors de la mise au point des matrices d'imputation et surtout lors de la détermination des valeurs initiales d'imputation dynamique. L'équipe de vérification doit spécifier les conditions dans lesquelles une valeur déterminée doit être affectée à une information laissée en blanc. Cette valeur devrait provenir d'une unité d'habitation précédente, dotée de caractéristiques voisines.

316. Tous les enregistrements concernant la population devraient comporter un numéro de série pour faciliter le traitement informatique. Les vérifications structurelles décrites au chapitre III vérifient la correspondance entre le numéro séquentiel et l'ordre des numéros de série.

317. L'équipe de vérification devrait contrôler uniquement les rubriques appropriées des enregistrements concernant la population. Or, le choix de ces rubriques peut différer en fonction de différentes caractéristiques (notamment milieu urbain/rural, climat, etc.). Bien qu'il y ait intérêt à effectuer des vérifications sélectives en fonction de

ces caractéristiques, peu de pays disposent du temps ou des compétences nécessaires pour définir et utiliser plusieurs tableaux de valeurs permettant de remplacer les données manquantes ou incohérentes. Les pays effectivement en mesure de suivre cette procédure supplémentaire sont encore moins nombreux.

318. Parfois, l'information recueillie sur le questionnaire ne s'applique qu'à certains groupes particuliers de population. Par exemple, les questions concernant la fécondité sont posées uniquement aux femmes et celles concernant l'activité économique, aux adultes.

319. Parfois, l'équipe de vérification doit autoriser pour certaines rubriques la mention « non indiqué ». De plus, elle ne dispose pas nécessairement de données suffisamment probantes pour imputer des réponses concernant certaines caractéristiques. La décision de laisser la mention « non indiqué » doit être soigneusement pesée, compte tenu de la nécessité d'obtenir des données tabulaires appropriées à des fins de planification et pour les besoins de l'action des pouvoirs publics. Dans la mesure où les données « non indiquées » présentent une distribution analogue à celle des données indiquées, la répartition des mentions « non indiqué », lorsque les planificateurs ont besoin de certaines informations, ne devrait pas poser de problèmes. En revanche, si la distribution des réponses « non indiqué » comportent un certain biais, l'application des procédures d'imputation après compilation risque d'être problématique, en particulier dans des zones peu étendues ou pour des caractéristiques bien précises. Par exemple, si des adolescentes interrogées refusent de répondre aux questions concernant leur fécondité et si aucune donnée n'est ainsi recueillie, le processus de vérification ne pourra aucunement faciliter l'obtention de cette information.

320. Les vérifications concernant les données sur la population sont d'ordinaire plus complexes que celles concernant l'habitation, en raison du caractère généralement beaucoup plus complexe des distributions croisées. La plupart des pays établissent chaque caractéristique d'habitation uniquement selon différents niveaux géographiques, alors qu'ils disposent parfois d'un grand nombre de strates de distribution croisée pour les données sur la population. Tel qu'indiqué plus haut, les pays qui décident de ne pas utiliser la méthode d'imputation dynamique devraient définir un identificateur pour les mentions « inconnu » à utiliser en présence de réponses non valides ou incohérentes.

321. Pour les pays qui appliquent une méthode d'imputation dynamique, les équipes de vérification devraient établir des matrices d'imputation simples dont les dimensions permettent de distinguer différentes caractéristiques démographiques. Dans la plupart des cas, le groupe d'âge et le sexe sont les variables primaires les plus indiquées pour utiliser la méthode d'imputation dynamique et elles devraient faire l'objet de la vérification initiale. Les bureaux nationaux de recensement/statistique qui effectuent des vérifications à plusieurs variables devraient prendre en compte simultanément l'âge, le sexe et différentes variables telles que le lien avec le chef de ménage et l'état matrimonial. Parmi les autres rubriques susceptibles de faciliter l'application d'une méthode d'imputation dynamique figurent le niveau d'instruction et la situation en matière d'emploi.

322. Les équipes de vérification doivent veiller tout particulièrement à ne pas introduire de distorsions dans les données au cours du processus d'imputation. Ainsi, il n'y a pas lieu de supposer que les données imputées et non imputées ont nécessairement des distributions identiques. Souvent, les données inconnues comportent une distorsion intrinsèque. Par exemple, les personnes relativement âgées sont moins disposées à déclarer leur âge que les personnes plus jeunes.

## A. Caractéristiques démographiques

323. Dans tout recensement, les données individuelles recueillies sur le lien avec le chef de ménage, le sexe, l'âge et l'état matrimonial sont fondamentales et devraient sans doute être vérifiées conjointement. Les répartitions par âge et par sexe des populations ou des sous-populations sont indispensables à pratiquement toutes les activités de planification fondées sur des recensements. Ces rubriques sont en outre nécessaires à l'obtention de tableaux de données significatives puisque pratiquement toutes les autres analyses reposent sur des distributions d'autres variables croisées par âge et par sexe.

324. La méthode dite à variables multiples (Fellegi-Holt) de vérification des données concernant la population et l'habitation a été présentée au chapitre II de ce manuel. Comme les variables démographiques font partie intégrante de l'organisation de tous les recensements, cette méthode devrait être utilisée dans la mesure où les délais et les compétences disponibles le permettent. De fait, la qualité de l'ensemble du jeu de données bénéficiera très certainement d'une vérification prioritaire axée sur les variables d'âge et de sexe et sur différentes variables particulières, afin de déterminer les erreurs ou incohérences. Les rubriques les plus fréquemment entachées d'erreurs sont vérifiées en premier, puis celles pour lesquelles les erreurs ou les incohérences sont moins fréquentes.

### 1. Lien avec le chef de ménage

325. La rubrique « lien avec le chef de ménage » facilite la détermination de la structure du ménage et de la famille. Elle figure pratiquement au début de la plupart des questionnaires de recensement et d'enquête et permet d'assurer le décompte de chaque individu de l'unité d'habitation. Recenseur et recensé utilisent les informations concernant les liens des différents membres du ménage avec le chef de ménage pour s'assurer que personne n'est oublié. La question de la détermination d'un chef de ménage unique et d'un conjoint unique (dans les sociétés non polygames) est traitée lors des vérifications structurelles.

#### a) Vérifications du lien avec le chef de ménage

326. Vu l'importance croissante des statistiques sur le lien avec le chef de ménage, il convient de définir soigneusement des vérifications qui permettent d'utiliser différents types de groupements familiaux et sous-familiaux pour divers types de tabulations. L'élaboration de codes appropriés de lien avec le chef de ménage facilitera certainement cette tâche (voir informations sur le type de famille, le numéro de sous-famille et les personnes apparentées au sein d'une sous-famille à l'annexe I).

327. Lorsqu'il est impossible d'attribuer une valeur à la variable lien avec le chef de ménage et si l'on n'utilise pas la méthode d'imputation dynamique, il faut alors attribuer la valeur « inconnu » aux réponses non valides ou incohérentes. Si par contre, la méthode d'imputation dynamique est utilisée, il est possible d'attribuer une valeur à la variable lien avec le chef de ménage d'après une matrice d'imputation définie en fonction de l'âge, du sexe ou d'autres caractéristiques appropriées. Les matrices d'imputation ne doivent pas imputer de liens avec le chef de ménage qui seraient incompatibles avec les liens déjà définis. Ainsi, il n'y a pas lieu d'imputer des deuxièmes et troisièmes conjoints, même dans les ménages polygames, sauf décision spécifiquement prise à cet effet par le groupe de vérification.

**b) Cas dans lesquels le chef de ménage doit figurer en premier**

328. Lorsque le chef de ménage ne figure pas en tant que première personne, les vérifications structurelles décrites au chapitre III signalent la possibilité d'utiliser un pointeur pour repérer dans la liste la position du chef de ménage. Si l'équipe de vérification souhaite que le chef de ménage soit la première personne de la liste, il est possible de le placer en première position, soit en modifiant le classement des personnes, soit en conservant la position du chef de ménage, mais en modifiant les liens avec celui-ci, tel qu'indiqué dans le chapitre consacré aux vérifications structurelles. La première méthode exige des compétences éprouvées en matière de programmation, tandis que la seconde risque d'altérer la qualité du jeu de données si cette opération n'est pas effectuée avec le plus grand soin.

**c) Cas dans lesquels les liens avec le chef de ménage sont codés « à l'envers »**

329. Les agents de recensement notent parfois « à l'envers » la variable « lien avec le chef de ménage » : au lieu d'enregistrer le lien de chaque personne avec le chef de ménage, ils enregistrent le lien du chef de ménage avec chacune des personnes qui le constituent. Ainsi, le lien de la troisième personne avec le chef de ménage prend la valeur « parent » et non « enfant ». On peut obtenir en définitive un ménage comportant quatre parents et non quatre enfants. Lorsque l'équipe de vérification se trouve confrontée à un problème systématique de ce type, il lui faut trouver une solution qui n'altère pas trop la qualité des données concernant le ménage.

330. La procédure employée pour inverser les liens avec le chef de ménage nécessite l'utilisation d'un fichier « contrôle visuel » contenant les liens originaux et les liens inversés et tenant compte du sexe des personnes recensées.

**d) Présence de plusieurs conjoints (polygamie)**

331. Effectuées conformément aux indications du chapitre III, les vérifications structurelles auront déjà vérifié que les ménages monogames comportent un chef de ménage et un seul, ainsi qu'un conjoint et un seul. Pour les ménages polygames, l'équipe de vérification doit décider si la prise en compte de ce type de lien avec le chef de ménage est autorisée ou non. Parfois, des ménages comportant apparemment des liens de polygamie avec le chef de ménage correspondent en réalité à des erreurs.

332. Par exemple, un chef de ménage et un conjoint peuvent être identifiés à l'intérieur d'un ménage, mais les personnes formant un autre couple se sont déclarées mutuellement en tant que conjoints, portant ainsi à trois le nombre total de conjoints. Le programme de vérification doit vérifier que le deuxième couple n'est pas formé en réalité du père et de la mère ou du fils et de la belle-fille, ou encore de la sœur et du beau-frère, ou d'une autre combinaison. Parfois, les relations de ce type peuvent être déterminées avec quelque certitude et parfois non. Sous réserve du codage des « liens avec le chef de ménage » présentés ci-dessus de façon détaillée, l'équipe de vérification peut escompter des imputations appropriées. Lorsque les conjoints supplémentaires sont effectivement des conjoints (dans des ménages polygames), alors, la vérification doit contrôler le sexe et, le cas échéant, l'âge des personnes en question.

**e) Présence de plusieurs parents**

333. Les ménages ne doivent pas compter plus de deux « parents » déclarés, et ces derniers ne doivent pas être du même sexe. Lorsque les parents déclarés sont plus

de deux, la valeur « autre personne apparentée » devrait être sans doute attribuée aux parents supplémentaires. Les recensements ou les enquêtes ont parfois défini un code correspondant aux parents ou aux parents par alliance (beaux-parents), ce qui permettrait de dénombrer jusqu'à quatre « parents » et non deux, à raison de deux parents ou plus par sexe.

f) Cas dans lesquels les recensements recueillent des données de lien avec le chef de ménage définissant le sexe

334. Certains recensements ou certaines enquêtes recueillent des informations particulières à un sexe concernant le lien avec le chef de ménage : par exemple, la qualité de « mari » ou de « femme » au lieu de celle de conjoint; ou la mention de « fils » ou de « fille » au lieu de celle d'« enfant ». Si ces réponses ne font pas l'objet d'une vérification, les tableaux de résultats risquent de contenir des données indiquant des « filles » de sexe masculin ou des maris de sexe féminin. L'équipe de vérification doit donc fixer l'ordre de priorité des contrôles, à savoir décider si le lien avec le chef de ménage ou le sexe doit prévaloir. Dans certains cas, tels que celui du mari ou de la femme, la vérification est plus importante que dans d'autres, tels que celui d'un jeune enfant. Il convient de noter qu'il n'est pas judicieux d'utiliser des liens définissant le sexe car les répétitions ne clarifient pas mais ne font qu'obscurcir la nature des liens avec le chef de ménage et nécessitent des contrôles supplémentaires.

g) Défaut de concordance entre le lien avec le chef de ménage et l'état matrimonial

335. Le lien avec le chef de ménage et l'état matrimonial doivent concorder lorsqu'ils se recoupent : les personnes qui déclarent le lien de « conjoint » doivent être « mariées » selon la rubrique « état matrimonial ». L'équipe de vérification choisit les variables à modifier lorsqu'il n'y a pas concordance. Parfois, compte tenu de l'ambiguïté des liens avec le chef de ménage, il faut définir avec le plus grand soin les spécifications de vérification. Par exemple, dans de nombreux pays, un beau-frère est aussi bien le frère d'un conjoint (et n'est pas alors nécessairement marié) que le conjoint d'un frère ou d'une sœur (et devrait alors être marié).

336. Les valeurs déclarées pour la variable « lien avec le chef de ménage » font actuellement apparaître plusieurs autres problèmes plus contemporains. Lorsque deux personnes non mariées, de sexes opposés cohabitent sans être mariées, la variable « lien avec le chef de ménage » pourrait être mise à la valeur « partenaire non marié » ou « conjoint ». Si le recensement ou l'enquête a prévu un code pour les partenaires non mariés, alors, l'état matrimonial ne devrait pas être mis à la valeur « marié », sauf si la personne est mariée à une personne différente de celle avec laquelle elle vit.

337. De même, lorsque, à notre époque, des personnes du même sexe cohabitent, elles peuvent être liées ou non par une relation sentimentale : dans le deuxième cas, elles peuvent être codées en tant que « colocataires » ou « camarades de chambre » ou « personnes non apparentées »; dans le premier cas, la catégorie « partenaire non marié » est sans doute appropriée dans certains pays. Ensuite, il appartient à l'équipe de vérification de décider de l'état matrimonial à attribuer. Les recensements ne peuvent faire de distinction entre relations platoniques et relations entre partenaires sexuels.

## 2. Sexe

338. L'information sur le sexe est sans doute l'une des plus faciles à recueillir, mais sa vérification mérite réflexion. Elle constitue l'une des variables les plus impor-

tantes, car la plupart des caractéristiques démographiques sont analysées par sexe. L'imputation d'une valeur à la variable « sexe » doit donner lieu à différentes comparaisons avec d'autres variables. Dans certains cas, l'attribution du sexe doit reposer sur des différences entre les sexes de personnes apparentées, généralement le chef de ménage et le conjoint, mais aussi entre parent et parents du conjoint. Les réponses « non valide » ou « inconnu » ne doivent généralement pas être laissées pour la variable « sexe » compte tenu de son importance particulière. Aussi convient-il de réfléchir soigneusement à la meilleure façon d'obtenir des résultats qui reflètent effectivement la situation réelle d'un pays. Un exemple d'organigramme concernant la vérification du sexe du chef de ménage et du conjoint est présenté à l'annexe IV.

339. Lorsque l'on a affaire à une personne qui n'est ni le chef de ménage ni son conjoint, on ne dispose d'aucune personne de référence; il y a lieu alors de vérifier d'autres rubriques de son enregistrement : s'il contient des données sur la fécondité, il y a lieu d'attribuer alors le code sexe féminin. Toutefois, si la réponse concernant le sexe de cette personne est manquante ou non valide, par exemple, mais s'il existe un conjoint dont le sexe est indiqué, alors le programme de vérification peut imputer à cette personne le sexe opposé.

a) Cas dans lequel le code du sexe est valide  
mais où le chef de ménage et le conjoint sont du même sexe

340. Lorsque les données font apparaître des indications fortement contradictoires, il y a lieu de modifier le code attribué au sexe, même si celui-ci est valide; l'enregistrement considéré mentionne par exemple la présence d'un deuxième couple marié, alors que le ménage compte déjà un chef de ménage et un conjoint. Si les deux personnes du deuxième couple déclarent le même sexe, il est alors possible d'utiliser les données sur la fécondité, notamment pour déterminer le sexe de chacune. Ensuite, l'enregistrement erroné peut être modifié.

b) Cas dans lequel une donnée sur la fécondité est associée à un homme  
ou ne figure pas en ce qui concerne une femme adulte

341. Le programme de vérification peut détecter des données sur la fécondité et/ou une indication d'enfants à la maison concernant un homme, erreur susceptible d'être liée au numéro de classement de la mère sur la liste ou à une variable similaire. En l'absence de conjoint, il est possible d'attribuer à la variable sexe la valeur sexe féminin, plutôt que de supprimer les données sur la fécondité. De la même manière, une personne adulte du sexe féminin à laquelle aucune information de fécondité ni aucun enfant n'est associé peut, dans certaines circonstances définies par l'équipe de vérification, faire l'objet d'une modification de la variable sexe à laquelle est attribuée la valeur sexe féminin.

c) Code du sexe non valide et présence d'un conjoint

342. Si la valeur de la variable sexe est laissée en blanc ou n'est pas valide, le programme de vérification doit utiliser les réponses recueillies pour le lien avec le chef de ménage et pour le sexe du conjoint, si le sexe indiqué pour celui-ci est valide, afin de déterminer le code correct à attribuer. Si la variable figurant pour le lien avec le chef de ménage est « chef de ménage », le programme vérifie ensuite la présence éventuelle d'un conjoint (en vérifiant s'il existe à l'intérieur du ménage une personne dont la variable « lien avec le chef de ménage » a la valeur « conjoint »). Une fois déterminé



le code attribué au sexe du conjoint, le code correspondant au sexe opposé est alors attribué au chef de ménage.

d) Non-validité du code du sexe du conjoint

343. Si la personne considérée est liée au chef de ménage en tant que « conjoint » et si le sexe du chef de ménage est indiqué, le programme attribue alors à cette personne le sexe opposé à celui du chef de ménage.

e) Code du sexe non valide et présence d'indications quant au sexe féminin

344. Il existe dans le questionnaire de nombreux indices permettant d'établir si la personne interrogée est une femme. Si le programme n'a pas encore déterminé le sexe de la personne en question et en présence d'indices permettant d'établir qu'il s'agit d'une femme, alors il convient d'attribuer le sexe féminin à l'enregistrement concernant cette personne. Par exemple, une telle modification est possible lorsque la vérification détecte une ou plusieurs données sur la fécondité. Parmi les données sur la fécondité figurent le nombre d'enfants, le nombre d'enfants vivant dans ce ménage, le nombre d'enfants vivant dans un autre lieu, le nombre d'enfants décédés et le nombre d'enfants nés vivants au cours des 12 derniers mois. Par ailleurs, il pourrait également s'agir de la mère d'une autre personne du ménage, de telle sorte que le numéro de ligne de cette personne doit être égal au numéro de ligne d'un autre membre du ménage.

f) Code du sexe non valide, lorsqu'il s'agit du mari du conjoint

345. Si la personne considérée est le mari d'un autre membre du ménage (d'après une rubrique mentionnant le numéro de ligne du mari), il convient d'attribuer à la variable sexe la valeur sexe masculin.

g) Code du sexe non valide en l'absence d'indications suffisantes pour déterminer le sexe

346. Si l'équipe de vérification n'utilise pas la méthode d'imputation dynamique, il faut alors attribuer un code « sexe inconnu ». Malheureusement, cela signifie que tous les tableaux devraient comporter une colonne ou une rangée supplémentaire, ou encore une série de colonnes ou de rangées relatives aux personnes de sexe inconnu. Or, comme le sexe est une variable binaire, on ne peut lui attribuer qu'une des deux valeurs, en commençant par l'une ou l'autre, puis affecter la valeur du sexe opposé à la deuxième donnée non valide et continuer de cette façon.

h) Remarques concernant les rapports des variables imputées à la variable sexe

347. Le sexe féminin est susceptible d'être attribué plus souvent si l'on procède par imputation statique. Les femmes adultes étant les seules personnes auxquelles sont affectées des données sur la fécondité, leur choix s'écarte légèrement d'une sélection aléatoire. C'est pourquoi, faute d'informations suffisantes, une personne pour laquelle on ne dispose d'aucune information est plus vraisemblablement de sexe masculin que féminin. Il est donc important d'envisager l'établissement de matrices d'imputation qui tiennent compte des proportions globalement observées entre les sexes.

### 3. Date de naissance et âge

348. L'âge est l'une des caractéristiques dont l'obtention et la vérification sont particulièrement délicates. Toutefois, cette variable est sans doute la plus importante parce que pratiquement toutes les caractéristiques démographiques sont étudiées en fonction de l'âge. Sa vérification exige des comparaisons approfondies avec les autres variables. Dans la plupart des cas, l'âge imputé doit être déduit des différences enregistrées entre les âges des personnes apparentées. Si ce type d'imputation n'est pas possible, il convient alors d'utiliser d'autres caractéristiques figurant dans l'enregistrement concernant la personne considérée. La vérification exigerait probablement une série de matrices d'imputation, notamment de l'âge selon le sexe, de l'état matrimonial, du lien avec le chef de ménage et de la fréquentation scolaire; de la différence d'âge entre mari et femme; et de la différence d'âge entre chef de ménage et conjoint.

#### a) Âge et date de naissance

349. La vérification structurelle calcule l'âge à partir de la date de naissance. Tout d'abord, il y a néanmoins intérêt à examiner la différence entre l'âge et la date de naissance. Comme il est indiqué dans la publication *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* (Organisation des Nations Unies, 2008, par. 2.136), on peut obtenir les renseignements sur l'âge soit en demandant la date de naissance (année, mois et jour), soit en demandant directement l'âge au dernier anniversaire.

350. L'utilisation de la date de naissance fournit des renseignements plus précis et il convient de l'utiliser chaque fois que les conditions le permettent. Si l'on ne connaît ni le jour ni le mois exact de la naissance, on peut indiquer la saison de l'année. La question relative à la date de naissance est appropriée dans tous les cas où les recensés connaissent cette date, qu'ils utilisent le calendrier solaire ou un calendrier lunaire, ou qu'ils l'identifient selon la tradition populaire par association d'un nom à un phénomène cyclique régulier.

351. Il est toutefois très important que le recenseur et le recensé soient bien d'accord sur le type de calendrier auquel correspond la date de naissance. Si certains recensés risquent de donner une réponse correspondant à un calendrier différent de celui des autres recensés, il faut prévoir d'indiquer sur le questionnaire quel est le calendrier utilisé. Il n'est pas souhaitable que le recenseur tente de convertir la date d'un calendrier à un autre. Le mieux est que cette conversion soit effectuée au moment de la conversion informatisée des données (Nations Unies, 2008, par. 2.137).

352. Pour plusieurs raisons, la question directe relative à l'âge risque de livrer des renseignements moins précis. Même si toutes les réponses sont basées sur la même méthode de calcul de l'âge, le recensé peut ne pas bien comprendre s'il lui est demandé d'indiquer son âge au dernier anniversaire, au prochain anniversaire ou à l'anniversaire le plus proche. En outre, d'autres problèmes peuvent se poser : l'âge peut être arrondi au chiffre le plus proche se terminant par zéro ou par cinq; le caractère estimatif de l'âge indiqué n'est pas nécessairement identifié comme tel, et l'âge peut très bien être délibérément indiqué de façon inexacte (Nations Unies, 2008, par. 2.138)

353. Nombre de bureaux de recensement/statistique recueillent soit la date de naissance, soit l'âge mais pas les deux. Comme il est indiqué dans *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* (Organisation des Nations Unies, 2008), l'âge exprimé en nombre d'années révolues revêt une très grande importance : les vérifications l'utilisent fréquemment et l'âge constitue souvent une dimension des matrices d'imputation. Mais surtout, nombre de

politiques nationales reposent sur l'âge, de telle sorte qu'il faut s'employer à obtenir des données de qualité optimale. Toutefois, même dans des conditions idéales, l'âge n'est pas toujours indiqué. Il faut donc veiller à ce qu'il soit calculé correctement et à ce qu'il soit compatible avec les autres réponses de chacun des membres du ménage.

#### Encadré 4

##### Vérification et imputation de l'âge

La vérification et l'imputation de l'âge doivent se dérouler comme suit :

- ☞ a) Attribuer un âge lorsque l'âge est laissé en blanc;
- ☞ b) Contrôler l'âge minimal des personnes mariées;
- ☞ c) Contrôler l'âge minimal du chef de ménage;
- ☞ d) Contrôler l'âge minimal des parents; et
- ☞ e) Effectuer tout contrôle propre au pays.

#### b) Relation entre âge et date de naissance

354. Lors de la vérification structurelle, l'âge doit être calculé s'il n'a pas été recueilli parallèlement à la date de naissance. La vérification de l'âge dans le cadre des vérifications individuelles sera un bon test de cohérence interne des enregistrements et d'un enregistrement à l'autre; toutefois, une première étape consiste à calculer l'âge à partir de la date de naissance et de la date du recensement. Il importe de contrôler l'âge calculé d'après la date de naissance pour vérifier qu'il coïncide avec les valeurs limites associées à la date du recensement.

355. L'âge des enfants nés au cours de l'année du recensement, mais après la tenue de ce dernier sera calculé comme égal à  $-1$  et doit donc être corrigé. Il conviendra probablement d'omettre les enfants dont la date de naissance est postérieure à la date du recensement. Si toutefois, après examen, la date de naissance se révèle erronée, en raison du dénombrement ou du traitement des données, il convient d'utiliser d'autres variables pour obtenir une meilleure estimation de l'âge.

#### c) Cas dans lequel l'âge calculé dépasse la valeur limite supérieure fixée

356. Pour les recensements effectués en l'an 2000 et après, la plupart des pays choisiront d'enregistrer les quatre chiffres de l'année de naissance. Pour les recensements effectués autour de 2010, la fourchette acceptable ira de 1900 jusqu'à l'année du recensement. Bien que l'utilisation de trois chiffres soit suffisante pour les traitements informatiques, ce choix risque de dérouter aussi bien les agents de recensement que le personnel de bureau. Parfois, l'âge calculé dépasse la limite supérieure définie par le recensement et doit donc être ajusté. Si le recensement a lieu en 2010 et une personne déclare être née en 1860, l'âge calculé de 150 ans sera vraisemblablement en dehors des limites d'âge acceptables et devra donc être modifié.

#### d) Vérification de l'âge

357. Le programme de vérification doit s'assurer de la cohérence de l'âge déclaré de la personne avec l'âge déclaré de la mère, du père ou de l'enfant de la personne en question. La vérification doit établir l'existence d'une différence d'âge minimal entre l'âge du père ou de la mère et l'âge de l'enfant. En cas d'imputation de l'âge, les contrôles de cohérence doivent utiliser des données telles que le nombre d'années pas-

sées dans la circonscription administrative, la durée de séjour (au lieu de résidence) et la dernière année d'études (niveau d'instruction). Il faut faire toutes ces modifications avant de modifier l'âge ou d'effectuer une imputation.

358. La vérification doit commencer par un contrôle de validité. Si l'âge est valide, les spécialistes peuvent souhaiter vérifier la compatibilité de l'âge de la personne en question avec celui de sa mère (si la mère fait partie des membres du ménage) et avec l'âge des enfants (s'il s'agit d'une femme et si ses enfants font partie du ménage). En cas d'incompatibilité entre ces âges, l'âge de la personne doit être noté et modifié ultérieurement.

e) Vérification de l'âge en présence du chef de ménage et du conjoint

359. L'étape suivante vise à déterminer la présence éventuelle d'un conjoint. Si tel est le cas, la validité de l'âge du conjoint doit être vérifiée (âge d'au moins X années, selon l'âge minimal à la date du mariage défini par le pays). Si l'âge n'est pas conforme à cette disposition, et en cas d'imputation dynamique, le programme attribue une valeur obtenue par imputation spéciale calculée à partir de la différence d'âge entre le mari et la femme. Les différences d'âge fluctuant moins que les âges eux-mêmes, la matrice d'imputation du programme enregistrera la différence d'âge entre le mari et la femme (telle qu'elle ressort d'enregistrements précédents). Cette valeur est ajoutée ou soustraite de l'âge du conjoint de la personne considérée pour fournir un âge calculé.

360. Pour assurer la cohérence de cet âge calculé avec d'autres caractéristiques, la matrice d'imputation doit également utiliser des variables telles que l'état matrimonial, la durée de résidence et la dernière année d'étude accomplie. Lorsque ces variables ne sont pas prises en compte, il peut en résulter un âge calculé inférieur au nombre d'années vécues dans la localité par la personne ou inférieur au nombre d'années de scolarité impliqué. Par exemple, la matrice d'imputation peut indiquer un âge de 8 ans alors que la personne peut avoir déclaré avoir déjà passé 10 ans à l'adresse mentionnée. Sans utilisation des autres variables, lorsque le programme de vérification contrôle la durée de séjour dans la localité, une autre matrice d'imputation aura pour effet de substituer une valeur erronée à une valeur exacte de la durée de séjour dans la circonscription administrative.

f) Vérification de l'âge du chef de ménage  
en l'absence du conjoint, mais en présence d'un enfant

361. Lorsque la détermination de l'âge du chef de ménage ne peut se faire par comparaison avec l'âge du conjoint, le programme peut alors examiner le lien avec le chef de ménage. Si la variable « lien avec le chef de ménage » indique la valeur « chef de ménage », le programme de vérification peut rechercher dans les autres enregistrements relatifs au ménage considéré la présence éventuelle d'un fils ou d'une fille dont l'âge est réputé exact. Le programme vérifie l'âge du fils ou de la fille et détermine un âge calculé par imputation dynamique en fonction de la « différence d'âge » selon une technique semblable à celle décrite ci-dessus pour les époux. Comme dans le cas précédent, l'âge calculé tient compte de la durée de séjour et du niveau d'instruction. L'âge ainsi obtenu sera compatible avec ces variables et permettra d'éviter des erreurs évidentes en tenant compte des années de séjour dans la circonscription administrative et de la dernière année de scolarité.

g) Vérification de l'âge du chef de ménage  
en présence d'un parent de ce dernier

362. Lorsqu'une personne ne peut être classée dans l'une des catégories ci-dessus, le programme peut vérifier la présence d'un parent parmi les membres du ménage. Dans cette hypothèse, il est alors possible de calculer un âge au moyen d'une matrice d'imputation utilisant la différence d'âge. La différence d'âge entre un enfant et un parent peut être beaucoup plus grande qu'entre mari et femme. C'est pourquoi cette vérification n'a lieu qu'en cas d'échec de la méthode utilisant la différence d'âge entre époux. L'âge calculé doit tenir compte des caractéristiques d'instruction et de la plus haute année complète de scolarité, de la durée de résidence dans la circonscription administrative, de l'état matrimonial, de la fécondité et de l'activité économique. Le programme doit supposer qu'une personne a atteint l'âge minimal acceptable si il/elle a déjà été marié/e, a des enfants ou déclare une activité économique quelconque.

h) Vérification de l'âge du chef de ménage  
en présence d'un petit-enfant de ce dernier

363. Lorsqu'une personne ne peut être classée dans l'une des catégories ci-dessus, le programme peut vérifier la présence d'un petit-enfant parmi les membres du ménage. Si cette présence est constatée, il est alors possible de calculer un âge au moyen d'une matrice d'imputation utilisant la différence d'âge. La différence d'âge entre le chef de ménage et le petit-enfant varie beaucoup plus qu'entre époux ou entre chef de ménage et enfant. C'est pourquoi cette vérification n'a lieu qu'en cas d'échec de la méthode utilisant la différence d'âge entre époux ou entre chef de ménage et enfant. L'âge calculé doit tenir compte des caractéristiques d'instruction et de la plus haute année complète de scolarité, de la durée de résidence dans la circonscription administrative, de l'état matrimonial, de la fécondité et de l'activité économique. Le programme doit supposer qu'une personne a atteint l'âge minimal acceptable si il/elle a déjà été marié/e, a des enfants ou déclare une activité économique quelconque.

i) Vérification de l'âge du chef de ménage  
en l'absence de toute autre donnée d'âge

364. Lorsqu'une personne ne peut être classée dans l'une des catégories décrites ci-dessus, le programme peut vérifier la présence parmi les membres du ménage d'une autre personne apparentée ou non au chef de ménage. Si une telle personne existe, et si son âge est indiqué, l'équipe de vérification doit décider d'utiliser toute information disponible au moyen d'une matrice d'imputation tenant compte de la différence d'âge. Toutefois, les différences d'âge entre le chef de ménage et les autres personnes apparentées ou non présentent de telles variations que l'équipe de vérification peut très bien décider de renoncer purement et simplement et de se contenter d'utiliser d'autres variables pour procéder à l'imputation dynamique de l'âge du chef de ménage. De toute façon, cette vérification a lieu uniquement après échec des méthodes utilisant la différence d'âge entre mari et femme, chef de ménage et enfant, chef de ménage et parent et chef de ménage et petit-enfant. Quelles que soient les modalités de calcul de l'âge, le résultat doit tenir compte des caractéristiques d'instruction et de la plus haute année complète de scolarité, de la durée de résidence dans la circonscription administrative, de l'état matrimonial, de la fécondité et de l'activité économique. Le programme doit supposer qu'une personne a atteint l'âge minimal acceptable si il/elle a déjà été marié/e, a des enfants ou déclare une activité économique quelconque.

j) Vérification de l'âge du conjoint  
après détermination de l'âge du chef de ménage

365. La vérification de l'âge du conjoint a lieu généralement en même temps que celle de l'âge du chef de ménage, car la vérification de l'âge de chacun utilise des informations concernant l'autre. Si toutefois la vérification a lieu séparément, lorsque l'âge du conjoint est non valide ou incohérente par rapport à d'autres variables, la meilleure estimation de l'âge du conjoint doit se faire à l'aide d'une matrice d'imputation dynamique utilisant la différence d'âge avec le chef de ménage, entre autres variables. Comme dans les cas précédents, l'âge calculé doit tenir compte des caractéristiques d'instruction et de la plus haute année complète de scolarité, de la durée de résidence dans la circonscription administrative, de l'état matrimonial, de la fécondité et de l'activité économique. Le programme doit supposer qu'une personne a atteint l'âge minimal acceptable si il/elle a déjà été marié/e, a des enfants ou déclare une activité économique quelconque.

k) Vérification de l'âge des autres couples mariés du ménage,  
lorsque l'âge de l'une des personnes est connu

366. La vérification doit d'abord établir si l'enregistrement considéré est celui d'une personne mariée. Si tel est le cas, le programme peut alors rechercher dans les autres enregistrements du ménage le conjoint de la personne en question. S'il ne trouve aucun conjoint, le programme passe à la deuxième partie de la vérification. S'il trouve un conjoint, il doit alors vérifier la validité de son âge (âge d'au moins X années, selon l'âge minimal à la date du mariage défini par le pays considéré). Si l'âge n'est pas cohérent et si l'un procède par imputation dynamique, le programme attribue alors une valeur d'imputation spéciale calculée à partir de la différence d'âge entre mari et femme. Les fluctuations des différences d'âge étant moins fortes que celles des âges eux-mêmes, la matrice d'imputation du programme enregistre la différence d'âge entre les époux (d'après les enregistrements précédents). Cette valeur est ajoutée ou retranchée de l'âge du conjoint de la personne considérée pour obtenir un âge calculé.

367. Pour assurer la cohérence de l'âge calculé par rapport à d'autres caractéristiques, la matrice d'imputation doit également inclure certaines autres variables, telles que l'état matrimonial, la durée de résidence et la plus haute année de scolarité. L'exclusion de ces variables peut se traduire par un âge calculé inférieur au nombre d'années de séjour dans la localité ou au nombre d'années de scolarité impliqué.

l) Vérification de l'âge d'un enfant  
après détermination de l'âge du chef de ménage

368. S'il s'agit du fils ou de la fille du chef de ménage, il est possible d'obtenir un âge calculé à partir de l'âge du chef de ménage, de la différence d'âge, de la durée de séjour et du niveau d'instruction. Là encore, l'âge calculé doit tenir compte des caractéristiques d'instruction, y compris de la plus haute année complète de scolarité, du nombre d'années de séjour dans la circonscription administrative, de l'état matrimonial, de la fécondité et de l'activité économique. Le programme doit supposer que la personne a atteint l'âge minimal acceptable si il/elle a déjà été marié/e, a des enfants ou déclare une activité économique quelconque.

m) Vérification de l'âge d'un parent  
après détermination de l'âge du chef de ménage

369. S'il s'agit d'un parent du chef de ménage, il est possible d'obtenir un âge calculé à partir de l'âge du chef de ménage, de la différence d'âge, de la durée de séjour et du niveau d'instruction. Là encore, l'âge calculé doit tenir compte des caractéristiques d'instruction, y compris de la plus haute année complète de scolarité, du nombre d'années de séjour dans la circonscription administrative, de l'état matrimonial, de la fécondité et de l'activité économique. Le programme doit supposer que la personne a atteint l'âge minimal acceptable si il/elle a déjà été marié/e, a des enfants ou déclare une activité économique quelconque.

n) Vérification de l'âge d'un petit-enfant  
après détermination de l'âge du chef de ménage

370. S'il s'agit d'un petit-enfant du chef de ménage, il est possible d'obtenir un âge calculé à partir de l'âge du chef de ménage, de la différence d'âge, de la durée de séjour et du niveau d'instruction. Là encore, l'âge calculé doit tenir compte des caractéristiques d'instruction, y compris de la plus haute année complète de scolarité, du nombre d'années de séjour dans la circonscription administrative, de l'état matrimonial, de la fécondité et de l'activité économique. Le programme doit supposer que la personne a atteint l'âge minimal acceptable si il/elle a déjà été marié/e, a des enfants ou déclare une activité économique quelconque.

o) Vérification de l'âge de toutes les autres personnes

371. L'équipe de vérification doit définir des matrices d'imputation appropriées concernant les autres personnes du ménage, apparentées ou non. Les principes directeurs adoptés à cet effet dépendront du type particulier de recensement ou d'enquête et des caractéristiques socio-économiques du pays. Par exemple, une personne qui a déjà été mariée, qui a eu des enfants ou qui a participé à une activité économique est susceptible d'avoir au moins l'âge minimal défini par le pays. Sur la base de cette information, si l'on utilise une imputation dynamique, la valeur tirée de la matrice d'imputation ne doit pas être inférieure à l'âge minimal. De même, si une personne fréquente un établissement scolaire, possède une certaine instruction ou sait lire et écrire, mais n'a pas la qualité de chef de ménage, n'a jamais été mariée et n'exerce aucune activité économique, elle doit être classée dans un groupe dont l'âge est inférieur à l'âge minimal défini pour les adultes mais supérieur ou égal à l'âge minimal requis pour la scolarité. La valeur fournie par la matrice d'imputation peut alors être déterminée pour les personnes dont l'âge est inférieur à l'âge minimal de scolarisation. Bien qu'imparfaite, cette technique limite le champ des valeurs que peut prendre la matrice d'imputation.

#### 4. État matrimonial

372. Dans la publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* (par. 2.144 à 2.151), l'état matrimonial est défini comme la situation d'une personne au regard des lois ou coutumes de son pays concernant le mariage. Les catégories d'état matrimonial suivantes doivent figurer parmi celles à identifier : a) célibataire (personnes n'ayant jamais été mariées); b) personnes mariées; c) veufs et non remariés; d) personnes divorcées et non remariées; et e) personnes mariées mais séparées. Dans certains pays, la catégorie b peut nécessiter l'établissement d'une sous-catégorie pour les personnes

légalement mariées mais ne vivant pas encore maritalement. Dans tous les pays, la catégorie *e* doit comprendre à la fois les séparations de droit et de fait qui peuvent, le cas échéant, constituer deux sous-catégories distinctes. Bien que les couples séparés puissent être considérés comme toujours mariés (vu qu'ils ne sont pas libres de se remarier), ni l'une ni l'autre des sous-catégories de la catégorie *e* ne doit être incorporée à la catégorie *b*. Dans certains pays, il faut tenir compte des unions coutumières (auxquelles le droit coutumier confère un caractère légal et contraignant) et des unions extralégales, souvent désignées sous le nom d'unions de fait (consensuelles).

#### a) Vérification de l'état matrimonial

373. L'équipe de vérification doit décider de l'âge minimal approprié au premier mariage à prendre en compte pour le recensement ou l'enquête. L'âge minimal au premier mariage (un certain âge *X*) peut varier selon la région du pays considéré ou selon le groupe ethnique. Si, par exemple, les populations rurales se marient plus tôt que les populations urbaines, les règles de vérification doivent en tenir compte. Normalement, le bureau national de recensement/statistique détermine l'âge minimal d'un mariage avant de procéder au dénombrement, de telle sorte que la question est posée uniquement aux personnes plus âgées. Les personnes plus jeunes sont classées automatiquement dans la catégorie « personnes jamais mariées ». Par contre, si la question de l'état matrimonial est posée à tout le monde, l'équipe de vérification doit alors mettre au point une vérification concernant l'ensemble de la population.

#### b) Attribution de la variable état matrimonial sans imputation dynamique

374. Bien que les données sur l'état matrimonial ne doivent être mises en tableaux que pour les personnes âgées d'au moins *X* années, avec *X* comme âge minimal au premier mariage, les équipes de vérification doivent décider de l'opportunité et de l'importance des contrôles effectués. Si le pays utilise uniquement les mentions « non indiqué » ou « inconnu » pour les réponses non valides ou incohérentes, toute réponse inadéquate doit alors être remplacée par le code correspondant à la mention « non indiqué ». Si, pour des personnes âgées de moins de *X* années, la réponse « jamais mariée » ne figure pas, elle doit alors être imputée; comme les bureaux de statistique divulguent des échantillons de données au public, il importe que les rubriques telles qu'état matrimonial aient toujours des réponses.

#### c) Attribution de la variable état matrimonial avec imputation dynamique

375. En cas d'imputation dynamique, la vérification de l'état matrimonial doit consister : *a*) à imputer une valeur lorsqu'une donnée déborde le champ des valeurs possibles; et *b*) à vérifier la cohérence entre l'état matrimonial déclaré, le lien avec le chef de ménage et l'âge.

#### d) Le conjoint devrait être marié

376. Toutes les personnes auxquelles le code « conjoint » du chef de ménage devraient avoir le code « personne mariée » dans la rubrique « état matrimonial ».

#### e) Conjoint dans un couple marié

377. Si le numéro de ligne du conjoint de la personne A (personne B) est une variable, la personne A devrait être indiquée comme conjoint de la personne B. En outre, A et B doivent être l'un et l'autre mariés et de sexes opposés.



f) S'il a la qualité de conjoint, le chef de ménage doit être marié

378. Si aucune valeur n'a été attribuée à l'état matrimonial, mais si la valeur du lien avec le chef de ménage est « chef de ménage », le programme doit vérifier la présence éventuelle du conjoint (en vérifiant la variable « lien avec le chef de ménage » des autres membres). Dans l'affirmative, le programme attribue l'état matrimonial « marié » au chef de ménage.

g) Chef de ménage sans conjoint ni enfants

379. Si le conjoint ne figure pas parmi les membres du ménage et si la personne considérée est de sexe masculin avec des enfants présents, le programme impute l'état matrimonial en fonction de l'âge et de la présence d'enfants. Si le ménage ne comprend pas d'enfants, le programme peut alors imputer l'état matrimonial en fonction de l'âge et de l'absence d'enfants. Un individu de sexe masculin recensé en tant que chef de ménage mais dont la femme est absente du ménage en question est très vraisemblablement divorcé, séparé ou veuf.

h) Imputation en cas d'échec de toutes les autres tentatives

380. Pour les personnes dont les codes sont hors limites et auxquelles les tests ci-dessus ne permettent pas d'attribuer un code approprié, la vérification suivante doit porter sur l'âge. Si un âge valide inférieur à l'âge X est mentionné, l'état matrimonial « personne jamais mariée » doit être attribué. Dans tous les autres cas, il convient d'attribuer une valeur au moyen d'une matrice d'imputation. La matrice d'imputation devrait être définie par sexe et par âge (deux dimensions); par sexe, par âge et par lien avec le chef de ménage (trois dimensions); ou encore, par sexe, par âge, par lien avec le chef de ménage et par nombre d'enfants (quatre dimensions). Là encore, les équipes de vérification devant avoir déterminé l'ordre dans lequel s'effectuent les contrôles, il importe lors de la définition des matrices d'imputation de bien distinguer les rubriques qui ont été vérifiées de celles qui ne l'ont pas été. Si une seule vérification du sexe et du lien avec le chef de ménage a précédé celle de l'état matrimonial, alors, la matrice d'imputation doit autoriser la mention « inconnu » pour les autres rubriques.

i) Relation entre l'âge et l'état matrimonial des jeunes

381. Pour toutes les personnes qui déclarent un état matrimonial valide autre que « jamais marié », il y a lieu d'effectuer un contrôle de cohérence avec l'âge. Toutes les personnes qui ont déjà été mariées doivent avoir au moins X années, X étant l'âge minimal autorisé dans le pays considéré pour le mariage. Si cet âge est inférieur à X ou s'il est laissé en blanc, d'autres contrôles de cohérence doivent être effectués en fonction des valeurs d'autres variables pertinentes (telles que le nombre d'enfants ou l'activité économique). Si les valeurs attribuées à ces variables ne sont pas valides, il y a lieu d'attribuer la valeur « jamais marié » ou « célibataire » à l'état matrimonial; dans tous les autres cas, l'état matrimonial ne doit pas être modifié.

## 5. Âge au premier mariage

382. D'après la publication *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* (par. 2.192), la « date du premier mariage est le jour, le mois ou l'année où le premier mariage a eu lieu ». Dans les pays où la date du premier mariage est difficile à obtenir, il est recommandé de recueillir des renseignements sur l'âge au mariage ou sur la durée écoulée depuis le mariage

(durée du mariage). Il faut prendre en compte non seulement les premiers mariages contractuels et les unions de fait, mais également les mariages coutumiers et religieux. Pour les femmes qui sont veuves, séparées ou divorcées au moment du recensement, il faut obtenir la date de la dissolution du premier mariage, l'âge à la dissolution ou le nombre d'années écoulées depuis. Les renseignements éventuels sur la dissolution du premier mariage livrent les données nécessaires pour calculer indirectement, lors du traitement, la durée du premier mariage. Dans les pays où les déclarations sur la durée du mariage sont plus fiables que l'âge de la femme concernée, les tableaux sur le total des naissances en fonction de la durée du mariage donnent de meilleures estimations de la fécondité que celles que l'on peut calculer à partir des données relatives aux enfants nés vivants, ventilées selon l'âge de la femme. On peut obtenir les données sur la durée du mariage en soustrayant de l'âge du moment l'âge au mariage, ou directement en calculant le nombre d'années depuis que le mariage a eu lieu.

383. La date du premier mariage doit être mentionnée pour toutes les personnes déjà mariées. Le programme doit vérifier la correspondance suivante : aucune information ne doit figurer dans le cas des personnes qui n'ont jamais été mariées, mais une indication valide de jour, de mois et d'année doit figurer dans le cas des personnes qui l'ont déjà été. Les équipes de vérification sont tenues de décider de l'éventuelle nécessité de mentionner un jour et un mois valides. Les pays n'utilisant pas la méthode d'imputation dynamique peuvent attribuer la valeur « inconnu » au jour et au mois; les pays utilisant cette méthode peuvent imputer le jour et le mois lorsque ces données sont manquantes.

a) L'âge au mariage des personnes qui n'ont jamais été mariées doit être laissé en blanc

384. Les personnes qui n'ont jamais été mariées ne doivent pas déclarer d'âge au premier mariage. Lorsqu'une donnée valide est inscrite dans le cas d'une personne qui n'a jamais été mariée, l'équipe de vérification doit décider de modifier l'état matrimonial ou de laisser en blanc l'âge de la personne. S'il faut modifier l'état matrimonial, les pays utilisant uniquement la mention « non indiqué » inscriront ce code. Les pays utilisant la méthode d'imputation dynamique doivent généralement utiliser l'âge et le sexe pour obtenir une réponse adéquate à la question sur l'état matrimonial.

b) Une valeur doit être indiquée dans le cas des personnes qui ont déjà été mariées

385. Pour l'année du premier mariage, les pays qui n'utilisent pas la méthode d'imputation dynamique peuvent attribuer la mention « non indiqué » ou « inconnu ». Les pays qui utilisent l'imputation dynamique peuvent utiliser d'autres variables, telles que l'âge du conjoint ou la différence d'âge entre conjoints, le nombre d'enfants et les naissances survenues au cours de l'année écoulée, pour déterminer adéquatement l'année du premier mariage.

## 6. Fécondité : nombre d'enfants et nombre d'enfants vivants

386. Le « nombre d'enfants » est égal au nombre total d'enfants nés vivants, à l'exclusion par conséquent des enfants mort-nés, des fausses couches et des avortements. Parfois, les démographes utilisent l'expression « nombre de naissances vivantes », mais nous utiliserons ci-après l'expression « nombre d'enfants ».

387. La population sur laquelle il convient de recueillir des données pour chacun des thèmes figurant dans cette section est constituée des femmes de 15 ans et plus

(ou ayant atteint l'âge minimal acceptable éventuellement défini), indépendamment de leur état matrimonial et de certaines sous-catégories particulières, telles que celle des femmes mariées ou ayant été mariées. Les pays qui ne recueillent ou n'exploitent pas de données pour les femmes de 50 ans et plus doivent concentrer leurs efforts sur la collecte de données sur les femmes âgées de 15 à 50 ans; lorsqu'ils enquêtent sur la fécondité récente, certains pays peuvent juger opportun d'abaisser de plusieurs années la limite d'âge inférieure (Nations Unies, 2007, par. 2.170).

#### a) Informations recueillies sur la fécondité

388. Dans la publication *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* (2007) l'Organisation des Nations Unies recommande l'obtention de renseignements concernant trois informations touchant à la fécondité : nombre total d'enfants, date de naissance du dernier enfant né vivant et âge de la mère à la naissance du premier enfant né vivant. Les réponses concernant l'âge au mariage, la date du mariage ou sa durée permettent d'améliorer les estimations de la fécondité fondées sur le nombre total d'enfants. De plus, de nombreux pays poursuivent la collecte d'informations sur le nombre d'enfants, ce qui facilite en particulier les études rétrospectives de la fécondité.

389. Les recensements et les enquêtes recueillent auprès de toutes les femmes des données sur la fécondité en tenant compte d'un âge minimal défini par pays et parfois également un âge maximal.

#### b) Principes généraux applicables à la vérification des données sur la fécondité

390. Les femmes d'un âge inférieur à celui désigné comme âge minimal pour la fécondité et toutes les personnes du sexe masculin devraient faire l'objet d'une vérification, et toute information concernant leur fécondité devrait être effacée.

391. La vérification des données concernant la fécondité a pour but d'assurer leur cohérence les unes envers les autres et vis-à-vis des données sur l'âge :

- a) Le nombre total d'enfants nés vivants ne peut être supérieur à l'âge minimal défini par le pays multiplié par un certain facteur. Ce facteur est égal à 1 lorsque 1 naissance par an est affectée aux femmes et à 1,5 correspondant à un an et demi entre deux naissances successives, et ainsi de suite. Voir la section ci-dessous intitulée « âge à la première naissance » pour la vérification visant à déterminer la différence d'âge minimal entre la mère et l'aîné de ses enfants nés vivants;
- b) Le nombre total d'enfants nés vivants ne peut être supérieur à la somme du nombre d'enfants qui vivent dans l'unité d'habitation ou dans un autre lieu et de ceux qui sont décédés. Lorsque le nombre total est supérieur à la somme des éléments constitutifs, les équipes de vérification doivent décider des éléments prioritaires à prendre en compte de façon à pouvoir effectuer les ajustements nécessaires;
- c) Si des données sont recueillies à la fois sur les enfants vivants et les enfants décédés, le nombre total de ces enfants ne peut être supérieur au nombre total de naissances;
- d) Le nombre d'enfants ne peut être inférieur à la valeur indiquée sous la rubrique « naissances au cours des 12 derniers mois »;

- e) Selon le pays et le nombre réel d'enfants, ainsi que le nombre d'enfants vivants, il est possible d'utiliser une matrice d'imputation pour le nombre de naissances au cours des 12 derniers mois, afin de pouvoir attribuer une réponse en fonction de l'âge et du nombre d'enfants. Toutefois, il faut faire preuve du plus grand soin dans l'attribution d'une valeur au nombre de naissances des 12 derniers mois lorsque cette rubrique est laissée en blanc. Pour la plupart des pays, un blanc signifie l'absence de naissances. L'attribution de valeurs particulières risque de fausser les données;
- f) Les pays recueillent parfois le nombre d'enfants, le nombre d'enfants vivants et différentes données sur la fécondité en les classant par sexe. En pareil cas, les vérifications figurant dans le présent document portent sur les agrégats, mais les pays peuvent souhaiter effectuer des contrôles supplémentaires pour tenir compte des autres informations disponibles. Ces contrôles supplémentaires consistent notamment à vérifier que le nombre d'enfants de sexe masculin est égal à la somme du nombre d'enfants vivants et d'enfants décédés de ce sexe et que le nombre d'enfants du sexe féminin est égal à la somme du nombre d'enfants de ce sexe vivants et décédés. Comme les vérifications portant sur le nombre d'enfants ne sont pas différenciées selon le sexe, il convient d'apporter des ajustements appropriés lorsque les totaux ne correspondent pas à la somme des parties.

#### c) Relation entre le nombre total d'enfants et le nombre d'enfants vivants

392. Les données sur le nombre d'enfants et le nombre d'enfants vivants permettent d'établir des estimations indirectes des taux de fécondité et de mortalité. Les résultats des recensements ou des enquêtes sont classés par cohortes de femmes d'une année ou de cinq années. Différents algorithmes produisent des estimations de taux de mortalité constants ou évolutifs. Toutefois, pour obtenir les meilleurs résultats possibles, il importe que les équipes de vérification déterminent avec soin la procédure appropriée de contrôle des données disponibles.

393. La difficulté à laquelle on se heurte lors de la conception d'une vérification générale vient du fait que les différents pays n'exigent pas les mêmes informations :

- a) Nombre d'enfants seulement;
- b) Nombre d'enfants et nombre d'enfants vivants (sexes combinés ou séparés);
- c) Nombre d'enfants, nombre d'enfants vivants, nombre d'enfants décédés (sexes combinés ou séparés);
- d) Nombre d'enfants, nombre d'enfants vivants à la maison, nombre d'enfants vivant ailleurs et nombre d'enfants décédés (sexes combinés ou séparés).

#### d) Vérification lorsque seul le nombre d'enfants est indiqué

394. Si le pays n'utilise pas la méthode d'imputation dynamique, il convient d'attribuer la valeur « inconnu » lorsque le nombre d'enfants indiqué est non valide ou manquant. Dans les pays qui utilisent la méthode d'imputation dynamique, les spécialistes doivent décider d'appliquer ou non ladite méthode à tous les éléments d'information. Dans ce dernier cas, il est possible d'obtenir le nombre d'enfants d'après le seul âge de la femme et au moins une autre caractéristique. L'emploi d'un tableau unidimensionnel de l'âge de la mère offre une autre possibilité. Parmi les autres caractéristiques à utiliser figurent des informations telles que le niveau d'instruction ou la religion, étant entendu que, dans nombre de pays, les différences à cet égard se traduisent par des taux de fécondité différents.

e) Vérification en cas d'indication du nombre d'enfants et du nombre d'enfants vivants

395. Si des réponses sont données à la fois pour les rubriques concernant le nombre d'enfants et le nombre d'enfants vivants, le programme de vérification doit établir les points suivants :

- a) Cohérence entre les informations fournies (le nombre d'enfants est-il supérieur ou égal au nombre d'enfants vivants ?);
- b) Compatibilité de chaque élément d'information avec l'âge de la femme;
- c) Compatibilité du nombre d'enfants avec le nombre d'enfants nés au cours de l'année écoulée (ou dernière naissance) si cette information est recueillie.

396. Les données concernant le nombre d'enfants et le nombre d'enfants vivants permettent aux démographes d'établir des estimations indirectes des taux de mortalité. Aussi les vérifications doivent-elles conserver le lien entre ces deux rubriques. Parfois, l'une des deux réponses est fournie, l'autre inconnue. Une vérification simple consisterait à supposer l'absence de décès parmi les enfants et à considérer ces deux chiffres identiques. Toutefois, dans cette hypothèse, l'estimation indirecte du taux de mortalité ne tiendrait pas compte des nouveau-nés décédés après la naissance, sous-estimant ainsi le taux de mortalité et surestimant l'espérance de vie. Lorsque les cas de ce type constatés lors d'un recensement ou d'une enquête sont peu nombreux, la qualité des données n'est guère altérée. En revanche, si leur fréquence est relativement élevée, comme cela devrait être le cas dans les pays procédant par estimation indirecte, les répercussions pourraient être importantes. La figure 27 donne un exemple de ce type de situation.

Figure 27

**Exemple de ménage comportant des données sur la fécondité**

Personne	Lien avec le chef de ménage	Sexes	Âge	Nombre d'enfants	Nombre d'enfants vivants
1	Chef de ménage	1	60		
2	Conjoint	2	60	5	99
3	Fille	2	40	3	3
4	Petite-fille	2	20	1	1
5	Petite-fille	2	18	0	0
6	Petite-fille	2	1		

Note : 99 = Données manquantes ou non valides.

397. Dans cet exemple, le conjoint déclare 5 enfants, mais pour une raison quelconque, le nombre d'enfants vivants n'a pas été indiqué. La personne interrogée ou l'agent recenseur n'a pas inscrit la valeur en question, ou l'opérateur de saisie l'a mal introduite. Nombre de pays élaborent une vérification attribuant la valeur « 5 » au nombre d'enfants vivants, d'après le nombre total d'enfants. Toutefois, cette façon de procéder introduit une distorsion dans les données.

398. En fait, il n'y a pas lieu de modifier la valeur indiquée. Les pays qui n'utilisent pas la méthode d'imputation dynamique peuvent choisir de conserver la mention « inconnu »; bien entendu, ce choix introduit également une distorsion car il est ainsi convenu que les réponses « inconnu » et « connu » ont la même distribution dans les tableaux de résultats. Un pays qui a besoin de données concernant le nombre total d'enfants et le nombre d'enfants vivants pour établir des estimations indirectes du taux de mortalité a par ailleurs vraisemblablement des difficultés de présentation des

données. En l'occurrence, le maintien de valeurs inconnues dans le jeu de données risque de fausser l'analyse finale. En effet, les femmes dont le nombre total d'enfants ou le nombre d'enfants vivants est affecté de la mention « inconnu » ne peuvent servir à l'estimation d'un taux de mortalité estimé, car la différence entre leur nombre total d'enfants et leur nombre d'enfants vivants ne peut être déterminée.

399. Les pays qui procèdent par imputation dynamique devraient envisager de déterminer l'élément d'information manquant en s'appuyant au moins sur l'autre donnée sur la fécondité et sur l'âge de la femme. Les matrices d'imputation peuvent être mises à jour en présence d'informations valides concernant l'âge de la femme, le nombre d'enfants et le nombre d'enfants vivants et peuvent ensuite être utilisées lorsque l'élément d'information est manquant. Ainsi, en l'absence d'informations sur le nombre total d'enfants, la matrice d'imputation pourra utiliser l'âge de la femme et le nombre d'enfants vivants. Par ailleurs, en l'absence d'informations sur le nombre d'enfants vivants, la matrice d'imputation utilisera l'âge de la femme et le nombre total d'enfants.

400. Par ailleurs, lors de l'établissement des matrices d'imputation, il ne faut pas oublier qu'il importe que le nombre d'enfants et le nombre d'enfants vivants soient compatibles avec la différence d'âge entre la mère et l'aîné des enfants (si cette information est connue) et que le nombre total d'enfants doit être compatible avec l'âge de la mère.

401. Par exemple, le nombre d'enfants imputé et l'âge de la mère pourraient différer d'au moins 12. Dès lors, dans certains cas, une matrice d'imputation utilisant des cohortes de femmes de 5 ans imputerait presque certainement des informations incompatibles.

402. La matrice d'imputation correspondant à la figure 28 fait apparaître l'âge des femmes (en ligne) et le nombre d'enfants (en colonne). Le contenu de la matrice correspond aux valeurs imputées du nombre d'enfants vivants. Les réponses recueillies seront parfois appropriées et parfois ne le seront pas. Si le programme de vérification détecte une femme de 19 ans qui a 5 enfants, le nombre de 5 enfants vivants serait sans doute conforme au critère de différence d'âge (une différence d'âge de 15 d'après le nombre d'enfants vivants et l'âge déclaré). En revanche, pour une femme de 15 ans, un nombre d'enfants égal à 5 (pour une différence d'âge égale à 10) comme un nombre d'enfants vivants égal à 4 (différence d'âge égale à 11) seraient l'un et l'autre inacceptables.

Figure 28

**Valeurs initiales pour la détermination du nombre d'enfants vivants avec des données valides d'âge et de nombre d'enfants vivants**

Nombre d'enfants	Âge												
	15	16	17	18	19	20	21	22	23	24	25-29	30-34	35+
0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	0	0	0
2		2	2	2	2	2	2	2	2	2	1	1	1
3			3	3	3	3	3	3	3	3	2	2	2
4					4	4	4	4	4	4	3	3	3
5					5	5	5	5	5	5	4	4	4

403. La matrice d'imputation donne de meilleurs résultats lorsque les âges concernent des femmes jeunes. Alors, seules les différences d'âge valides concernant cet âge particulier sont inscrites dans la matrice d'imputation et seules les réponses valides peuvent en être extraites.

f) Vérification en cas de mention du nombre total d'enfants, du nombre d'enfants vivants et du nombre d'enfants décédés

404. Le « nombre total d'enfants » est égal à la somme du « nombre d'enfants vivants » et du « nombre d'enfants décédés ». Tout défaut de cohérence peut être corrigé comme il est indiqué ci-dessous.

i) Indication des trois informations

405. En présence des trois informations, le programme de vérification doit déterminer les points suivants :

- a) Cohérence interne des trois informations, le nombre d'enfants devant être égal à la somme du nombre d'enfants vivants et du nombre d'enfants décédés;
- b) Cohérence de chacune des trois informations avec l'âge de la femme;
- c) Cohérence du nombre d'enfants avec le nombre d'enfants nés au cours de l'année écoulée (ou depuis la dernière naissance), si cette information est recueillie.

406. Si les trois points ci-dessus sont établis, la vérification est terminée. Par contre, si l'un d'eux ne l'est pas, le programme de vérification doit corriger le défaut de cohérence correspondant. Les trois informations ne sont pas nécessairement compatibles entre elles. Par exemple, une femme peut avoir eu 5 enfants, mais ne déclarer que 2 enfants vivants et 2 décédés. L'équipe de vérification doit alors choisir la variable qui l'emportera sur les autres. Souvent, la femme se souvient vraisemblablement du nombre total d'enfants qu'elle a eus, mais peut oublier le nombre exact d'enfants décédés. L'équipe de vérification peut alors décider d'accepter le chiffre déclaré du nombre total d'enfants et celui du nombre d'enfants vivants et de soustraire ce dernier chiffre du total pour obtenir un chiffre cohérent pour le nombre d'enfants décédés.

ii) *Indication de deux informations*

407. Comme le nombre total d'enfants (NTE) est égal à la somme du nombre d'enfants vivants (EV) et d'enfants décédés (ED), si deux de ces trois informations sont connues, le programme informatique peut déterminer la troisième :

Si NTE et EV sont connus,  $ED = NTE - EV$

Si EV et ED sont connus,  $NTE = EV + ED$

Si NTE et ED sont connus,  $EV = NTE - ED$

Ces vérifications devraient normalement être faites en premier. Lorsque le programme a établi la validité et la cohérence des trois informations, la vérification est terminée.

iii) *Indication d'une information*

408. Quand une seule des trois informations est indiquée et le pays considéré n'utilise pas la méthode d'imputation dynamique, les deux autres rubriques doivent être affectée de la mention « inconnu ». Par contre, si le pays utilise la méthode d'imputation dynamique, les équipes de vérification doivent définir un moyen de déter-

miner au moins une autre information, la troisième pouvant ensuite être obtenue par soustraction ou addition. Une matrice à deux dimensions peut servir à déterminer la deuxième donnée sur la fécondité en s'appuyant sur la première et sur l'âge de la mère. Si le nombre d'enfants est connu, par exemple, le nombre d'enfants vivants peut être déduit à l'aide de la matrice d'imputation, comme il est indiqué ci-dessus, et l'on peut ensuite obtenir le nombre d'enfants décédés en procédant par soustraction. De même, si le nombre d'enfants vivants est connu, on peut obtenir le nombre total d'enfants à partir de la matrice d'imputation de l'âge de la mère, et l'on obtient ensuite le nombre d'enfants décédés par soustraction.

iv) *Absence de toute information*

409. Si aucune des trois informations n'est indiquée, l'équipe de vérification doit décider de la marche à suivre. Si le pays n'utilise pas la méthode d'imputation dynamique, les trois informations doivent être affectées de la mention « inconnu », et aucune ne doit être utilisée par les méthodes d'estimation indirecte des taux de mortalité ou de fécondité. Dans les pays utilisant la méthode d'imputation dynamique, c'est aux spécialistes de décider d'appliquer cette méthode aux trois informations.

410. Si les spécialistes décident d'utiliser l'imputation dynamique, le nombre d'enfants peut s'obtenir à partir de l'âge de la mère et d'au moins une autre caractéristique. On peut également utiliser un tableau de valeurs à une seule dimension (âge de la mère). Au nombre des autres caractéristiques pouvant être prises en compte figurent le niveau d'instruction et la religion.

411. Une fois la première information déterminée, on peut obtenir la deuxième donnée sur la fécondité en suivant les opérations décrites ci-dessus pour la vérification lorsque l'on ne dispose que d'une information. Ensuite, on peut déduire la troisième information des deux premières. Ces trois informations doivent être compatibles car la mise à jour des matrices d'imputation exige que toutes les données soient compatibles. Enfin, les données obtenues sur la fécondité doivent également être compatibles avec celles des autres femmes de la même zone géographique car les données concernant ces femmes sont utilisées pour la mise à jour de la matrice d'imputation.

g) *Vérification en cas d'indication du nombre d'enfants, du nombre d'enfants vivant à la maison, du nombre d'enfants vivant à l'extérieur et du nombre d'enfants décédés*

i) *Indication des quatre informations*

412. En présence des quatre informations, le programme de vérification doit déterminer les points suivants :

- a) Cohérence interne des quatre informations, le nombre d'enfants devant être égal à la somme du nombre d'enfants vivant au foyer familial et du nombre d'enfants décédés;
- b) Cohérence de chacune des quatre informations avec l'âge de la femme;
- c) Cohérence du nombre d'enfants avec le nombre d'enfants nés au cours de l'année écoulée (ou depuis la dernière naissance), si cette information est recueillie.

413. Si les quatre points ci-dessus sont établis, la vérification est terminée. Par contre, si l'un d'eux ne l'est pas, le programme de vérification doit corriger le défaut de cohérence correspondant. Comme dans le cas décrit ci-dessus d'indication de trois informations, les quatre données indiquées ne sont pas nécessairement compatibles



entre elles. Là encore, l'équipe de vérification doit choisir la variable qui l'emportera sur les autres. Souvent, la femme se souvient vraisemblablement du nombre total d'enfants qu'elle a eus, mais peut oublier certains de ceux qui ont quitté le foyer familial ou le nombre exact d'enfants décédés. L'équipe de vérification peut alors décider d'accepter le chiffre déclaré du nombre total d'enfants et celui du nombre d'enfants vivants (la somme des enfants vivant au foyer familial et à l'extérieur) et de soustraire ce chiffre du total pour obtenir des valeurs cohérentes pour les autres variables. L'équipe de vérification peut avoir à cet effet à concevoir des algorithmes décrivant les différentes combinaisons possibles d'événements.

ii) *Indication de trois des quatre informations*

414. Le nombre total d'enfants (NTE) étant égal à la somme du nombre d'enfants vivants au foyer familial (EVF), du nombre d'enfants vivant à l'extérieur (EVE) et du nombre d'enfants décédés (ED), si trois des quatre informations sont connues, le programme informatique peut déterminer la quatrième :

Si NTE, EVF et EVE sont connus,  $ED = NTE - EVF - EVE$

Si EVF, EVE et ED sont connus,  $NTE = EVF + EVE + ED$

Si NTE, EVF et ED sont connus,  $EVE = NTE - EVF - ED$

Si NTE, EVE et ED sont connus,  $EVF = NTE - EVE - ED$ .

iii) *Indication de deux informations sur quatre*

415. Si deux informations seulement sont connues, l'équipe de vérification doit décider de la marche à suivre. Par exemple, dans de nombreux pays, les femmes ne signalent pas le nombre d'enfants décédés. L'autre information particulièrement susceptible d'être omise est le nombre d'enfants vivant en dehors de l'unité d'habitation, donnée qui ne peut être obtenue directement. Il faut donc procéder soigneusement à l'élaboration du questionnaire, à l'exécution du dénombrement ainsi qu'au dépouillement pour obtenir la meilleure qualité possible de toutes les données sur la fécondité.

416. Le nombre d'enfants qui vivent au foyer familial considéré (EVF) peut s'obtenir en dénombrant les enfants recensés dans la même unité d'habitation. Dans la meure où une seule femme du foyer en question présente le lien correspondant avec le chef de ménage, un simple décompte doit indiquer le nombre d'enfants qui vivent dans cette unité. Lorsque plusieurs femmes présentent ce même type de lien avec le chef de ménage, le programme de vérification peut encore être appliqué, à condition que les enfants soient recensés immédiatement après leur mère lors de la collecte de données. En cas d'échec de toutes les tentatives, les pays qui procèdent par imputation dynamique pourraient attribuer le nombre d'enfants vivant dans l'unité d'habitation considérée d'après l'âge de la mère et d'après l'une des autres variables connues. (Voir les règles générales énoncées ci-dessous pour l'imputation de données individuelles sur la fécondité à partir des autres informations disponibles et de l'âge de la mère). Il importe d'utiliser l'âge de la femme chaque fois que cela est possible, ainsi que le nombre d'enfants vivant dans l'unité d'habitation, vivant en dehors ou décédés.

417. À titre d'exemple, supposons que les données relatives au nombre total d'enfants et au nombre d'enfants décédés soient valides et que celles concernant le nombre d'enfants vivant au foyer familial et le nombre d'enfants vivant à l'extérieur ne le soient pas. Dans ce cas, on peut déterminer le nombre d'enfants vivant au foyer familial en dénombrant les enfants présentant le lien approprié avec la mère (à supposer que la mère soit le chef de ménage). Trois des quatre informations sont alors connues, et l'on peut déterminer la quatrième, le nombre d'enfants vivant à l'extérieur, en effectuant la soustraction  $EVE = NTE - EVF - ED$ .

418. Toutefois, lorsque deux informations seulement sont connues, le nombre d'enfants ainsi que le nombre d'enfants vivant au foyer familial devront plus vraisemblablement faire l'objet d'un nouveau codage. D'ordinaire, les femmes signalent volontiers le nombre d'enfants qu'elles ont eus, tandis que l'on peut obtenir les renseignements concernant le nombre d'enfants vivant au foyer familial par observation ou auprès des personnes interrogées lors du dénombrement; or, ces solutions ne sont pas applicables aux enfants vivant à l'extérieur ou aux enfants décédés. L'équipe de vérification peut alors recourir à une matrice d'imputation à partir de l'âge de la mère et du nombre total d'enfants (NTE), ou mieux encore, à partir de l'âge de la mère, du nombre total d'enfants (NTE) et du nombre d'enfants vivant au foyer familial (EVF). Ces variables permettent de déduire les informations requises à partir des renseignements concernant une femme présentant les mêmes caractéristiques quant au nombre d'enfants vivant à l'extérieur (EVE).

419. Les pays utilisant uniquement la matrice à deux dimensions [âge de la mère et nombre total d'enfants (NTE)] sans inclure la troisième dimension [nombre d'enfants vivant au foyer familial (EVF)] risquent d'obtenir pour le nombre d'enfants vivant à l'extérieur une valeur incompatible avec les deux autres éléments d'information. Dans le cas d'une femme âgée de 25 ans dont le nombre total d'enfants NTE = 5, la matrice d'imputation peut indiquer un nombre d'enfants vivant à l'extérieur égal à 3. Si le nombre d'enfants vivant au foyer familial est de 2, la vérification ne pose pas de problème. La valeur à attribuer au nombre d'enfants décédés est de 0 et les données sur la fécondité doivent alors être les suivantes : NTE = 5; EVF = 2; EVE = 3 et ED = 0.

420. Toutefois, le nombre d'enfants vivant au foyer familial pourrait être en fait de 4, l'âge de la mère et le nombre total d'enfants étant les seules données utilisées pour déterminer le nombre d'enfants vivant à l'extérieur. Un nombre de 3 enfants vivant à l'extérieur entraînerait une incompatibilité entre les différentes informations. En effet, le nombre total d'enfants (5) serait alors inférieur à la somme du nombre d'enfants vivant au foyer familial (4) et à l'extérieur (3), soit un total de 7. Il conviendrait alors d'utiliser une matrice à trois dimensions : pour un nombre total d'enfants égal à 5 et 4 enfants vivant au foyer familial, la matrice d'imputation pourrait présenter une valeur égale à 1 pour le nombre d'enfants vivant à l'extérieur (et la valeur 0 serait obtenue par soustraction pour le nombre d'enfants décédés). Sinon, la valeur donnée par la matrice d'imputation devrait être de 0 pour le nombre d'enfants vivant à l'extérieur. Il faudrait établir des matrices d'imputation analogues pour les autres couples de valeurs connues, conformément à la figure 29.

Figure 29

## Exemple de matrices d'imputation à définir pour des couples de valeurs connues

Si ces valeurs sont connues ...		Utiliser l'imputation dynamique pour connaître l'une des valeurs suivantes (procéder ensuite par soustraction ou addition)	
Nombre total d'enfants	Enfants vivant au foyer familial	Enfants vivant à l'extérieur	Enfants décédés
Nombre total d'enfants	Enfants vivant à l'extérieur	Enfants vivant au foyer familial	Enfants décédés
Nombre total d'enfants	Enfants décédés	Enfants vivant au foyer familial	Enfants vivant à l'extérieur
Enfants vivant au foyer familial	Enfants vivant à l'extérieur	Nombre total d'enfants	Enfants décédés
Enfants vivant au foyer familial	Enfants décédés	Nombre total d'enfants	Enfants vivant à l'extérieur

Si ces valeurs sont connues ...		Utiliser l'imputation dynamique pour connaître l'une des valeurs suivantes (procéder ensuite par soustraction ou addition)	
Enfants vivant à l'extérieur	Enfants décédés	Nombre total d'enfants	Enfants vivant au foyer familial

421. Dans chaque cas, deux des quatre données sont connues. La troisième s'obtient par imputation dynamique et la quatrième par soustraction ou addition. Il appartient aux équipes de vérification de choisir la meilleure voie à suivre sur la base des conditions culturelles.

iv) *Indication d'une seule information*

422. Lorsqu'une seule des quatre informations est connue, la situation est encore plus problématique. Les pays doivent alors décider d'une procédure à suivre lorsque si peu d'informations sont disponibles. En cas de recours à la méthode d'imputation dynamique, comme il est indiqué plus haut, la première matrice d'imputation utiliserait une donnée telle que l'âge de la mère et la donnée connue pour créer une matrice à deux dimensions pour l'imputation de l'une des autres variables. Une fois que deux éléments sont déterminés, par définition, les deux autres sont encore inconnus. Ainsi, le fait de continuer à utiliser la méthode d'imputation dynamique pour la troisième donnée ne devrait pas créer d'incompatibilité avec les autres éléments puisqu'ils sont inconnus. Le processus décrit ci-dessus pour deux éléments connus et deux éléments inconnus est utilisé pour déterminer un troisième élément. Ensuite, le quatrième élément s'obtient par soustraction. Les quatre éléments d'information ainsi déterminés devraient être tous compatibles.

v) *Indication d'aucune information*

423. Lorsqu'aucune de ces quatre informations n'est indiquée, l'équipe de vérification doit décider de la marche à suivre en l'absence de toute donnée connue. Sans recours à la méthode d'imputation dynamique, tous les éléments doivent être affectés de la mention « inconnu », et aucun d'entre eux ne doit être utilisé pour des méthodes indirectes d'estimation du taux de mortalité ou de fécondité. Dans les pays qui utilisent l'imputation dynamique, les spécialistes doivent décider s'ils souhaitent appliquer cette méthode à tous les éléments d'information.

424. Si les spécialistes optent pour l'imputation dynamique, le nombre total d'enfants peut être déterminé d'après l'âge de la mère et au moins une autre caractéristique. Il est également possible d'utiliser un tableau à une dimension (portant uniquement l'âge de la mère). Parmi les autres caractéristiques à utiliser figurent le niveau d'instruction ou la religion, car on sait que, dans beaucoup de pays, les différences dans ces domaines influent sur les taux de fécondité.

425. Une fois la première information obtenue, la méthode employée ci-dessus quand on ne dispose que d'une donnée peut être utilisée pour déterminer la deuxième donnée sur la fécondité. La troisième donnée peut ensuite être déduite des deux premières et la quatrième, par soustraction. Les quatre éléments d'information devraient être compatibles car les matrices d'imputation ne doivent être mises à jour que lorsque toutes les données sont compatibles. Enfin, les données sur la fécondité ainsi obtenues doivent également être compatibles avec celles des autres femmes de la zone géographique, car ce sont les données concernant ces dernières qui servent à la mise à jour de la matrice d'imputation.

#### h) Cas particulier d'indication de cinq éléments d'information ou plus

426. À mesure que les migrations internationales prennent plus d'importance dans certains pays plus petits, des informations supplémentaires sont recueillies sur les enfants vivant hors du foyer familial. Lorsque la variable « enfants vivant à l'extérieur » se divise en « enfants vivant à l'extérieur mais dans le pays » et « enfants vivant à l'étranger », il faut étendre les procédures utilisées pour quatre variables — au foyer familial, à l'extérieur, décédés et total — de manière à tenir compte de ces informations supplémentaires. Par ailleurs, comme il a été indiqué, il est bon d'avoir un tableau à une dimension pour l'âge de la mère, avec données complètes sur la fécondité lorsque tous les éléments d'information sont valides, cohérents et compatibles avec cet âge; puis, lorsque les données sur la fécondité ne sont pas compatibles (notamment avec l'âge), on peut utiliser l'imputation en l'appliquant à l'ensemble du tableau.

#### i) Importance de l'utilisation d'un donneur unique pour toutes les données sur la fécondité

427. Il est donc très important, dans la mesure du possible, d'imputer toutes les informations à partir d'une seule femme lorsqu'aucune indication n'est fournie. Pour être sûr que toutes les informations proviennent de la même femme, il peut être nécessaire d'établir des matrices d'imputation qui utilisent toutes les données sur la fécondité. Dans ce cas, les matrices d'imputation ne peuvent être mises à jour que lorsque le programme de vérification a établi la concordance de toutes les données sur la fécondité. Comme il est indiqué au paragraphe 426 ci-dessus, il est préférable de ne pas imputer les données une par une, mais lorsque plusieurs données sont manquantes, d'utiliser l'ensemble des informations relatives à une même femme.

#### j) Relation entre les enfants du ménage, les enfants vivant à la maison et les enfants vivants

428. Lorsque les pays utilisent la méthode fondée sur le nombre total d'enfants pour aider à la vérification de la fécondité lors de sa mise au point et de son application, les données sur les enfants vivant au foyer familial et sur la matrice mère-enfant peuvent faciliter le contrôle de la fiabilité des résultats de la vérification. Comme très peu de pays l'utilisent pour faciliter la vérification, cette méthode reste expérimentale; cela dit, les résultats s'annoncent prometteurs.

### 7. Fécondité : date de naissance du dernier enfant né vivant et naissances survenues dans les 12 mois ayant précédé le recensement

429. Les informations sur les dernières naissances permettent d'établir des estimations de la fécondité juste avant le recensement. Une méthode consiste à obtenir la date de naissance (jour, mois et année) du dernier enfant né vivant ainsi que son sexe pour déterminer si cet enfant est vivant ou non. Une deuxième méthode consiste à recueillir des données sur les naissances survenues dans les 12 mois ayant précédé le recensement; il est plus facile pour les agents de recensement et les personnes interrogées d'utiliser cette seconde méthode car elle ne nécessite qu'un « oui » ou un « non » plutôt qu'une date exacte.

430. Pendant le traitement, on peut tirer une estimation du nombre d'enfants nés vivants au cours des 12 mois ayant précédé immédiatement la date du recensement des données sur la date de naissance du dernier enfant né vivant (puis la garder comme

nouveau code). Pour l'estimation des taux de fécondité liés à certains âges particuliers et de certaines autres mesures de la fécondité, les données fournies par cette méthode sont plus précises que les informations sur le nombre d'enfants auxquels une femme a donné naissance au cours des 12 mois ayant précédé immédiatement la date du recensement (Nations Unies, 2008, par. 2.188 à 2.191).

431. Il convient de noter que les informations relatives à la date de naissance du dernier enfant né vivant ne donnent pas le nombre total d'enfants nés vivants au cours de ladite période de 12 mois. Même si les données sur le dernier enfant né vivant ne comportaient pas d'erreurs, la question ne permettrait d'établir que le nombre de femmes qui ont eu au moins un enfant né vivant au cours de la dernière période de 12 mois et non le nombre des naissances puisqu'une faible proportion des femmes peuvent avoir eu plus d'un enfant au cours d'une même année (Nations Unies, 2007, par. 2.189).

432. On ne collectera de données que sur les femmes âgées de 15 à 50 ans qui ont indiqué avoir eu au moins un enfant vivant au cours de leur existence. De même, les renseignements doivent porter sur toutes les catégories de femmes, quel que soit leur état matrimonial, pour lesquelles on collecte des données sur les naissances totales ventilées selon le sexe. Si les données sur la descendance finale ne sont collectées que pour un échantillon de femmes, le même échantillon doit servir pour la collecte de données sur la fécondité du moment (Nations Unies, 2007, par. 2.190).

433. Le programme de vérification devrait porter sur les points suivants : la date de naissance du dernier enfant devrait être indiquée pour toutes les femmes d'un âge compris entre les âges minimaux et maximaux définis par le pays. Le programme devrait vérifier l'existence de cette correspondance. Par exemple, aucune information ne devrait figurer pour les hommes et les femmes n'appartenant pas au groupe d'âge ainsi défini. De plus, pour les femmes du groupe d'âge considéré et ayant au moins un enfant, le jour, le mois et l'année de la dernière naissance devraient être valides (ou encore, il devrait être précisé si la naissance a eu lieu dans les 12 derniers mois, si cette question est utilisée).

434. Il incombe à l'équipe de vérification de décider si le jour et le mois doivent être valides : les équipes de vérification utilisant la méthode d'imputation dynamique peuvent imputer le jour et le mois lorsque ces informations manquent; celles qui n'utilisent pas cette méthode doivent attribuer la mention « inconnu » à ces deux données. Si les spécialistes du sujet, généralement des démographes, veulent l'âge effectif des mères à la naissance de leurs enfants comme nouveau code pour l'analyse de la fécondité, il leur faut tout au moins imputer le mois de la dernière naissance s'il n'est pas indiqué. Ils pourraient alors obtenir le nouveau code.

435. De même, certains démographes veulent pouvoir analyser les mois écoulés depuis la dernière naissance. La vérification de l'année et du mois de la dernière naissance fournit les informations nécessaires pour déterminer le nombre de mois écoulés depuis cette naissance. Lorsque le jour du dernier mois est également recueilli, il peut servir à déterminer le nouveau code pour les mois écoulés depuis la dernière naissance. (Voir la méthode de détermination d'un nouveau code pour les mois écoulés depuis la dernière naissance à l'annexe I).

436. Si l'année de naissance du dernier enfant est manquante ou non valide, la mention « non indiqué » ou « inconnu » peut être attribuée lorsque le pays n'utilise pas la méthode d'imputation dynamique; par contre, les pays qui utilisent cette méthode peuvent déterminer la date de naissance du dernier enfant à partir d'autres variables telles que l'âge et le nombre d'enfants.

437. Vu l'importance de l'utilisation de la date de la dernière naissance comme indication des niveaux récents de fécondité aux échelons national, régional et local,

d'autres vérifications devraient être envisagées. Il pourrait être utile à cet égard de s'assurer de la présence d'un ou de plusieurs enfants de moins d'un an au sein d'un ménage, et d'utiliser les liens entre la mère et cet enfant ou ces enfants (ou le numéro de personne de la mère pour l'enfant, s'il a été recueilli) pour déterminer si l'enfant est déclaré comme la dernière naissance pour la mère. Ces vérifications devraient se faire dans les deux sens : les enfants de moins d'un an devraient être vérifiés par rapport aux mères et les dernières naissances devraient être vérifiées par rapport à la liste de composition du ménage.

438. Les pays qui collectent également des données sur les décès survenus durant l'année ayant précédé le recensement ou l'enquête peuvent décider de procéder à une vérification des décès d'enfants de moins d'un an au cours de l'année en question par rapport aux dernières naissances, lorsque le dernier enfant né est déclaré « décédé » ou « n'étant plus en vie ». Si cette vérification n'est pas possible au cas où la mère ne réside plus dans le logement pour cause de décès ou de départ, ou si, pour une raison quelconque, l'enfant n'est pas déclaré, une certaine proportion de décès d'enfants peut encore être vérifiée de cette façon.

## **8. Fécondité : âge à la première naissance**

439. L'âge de la mère à la première naissance vivante sert à calculer indirectement la fécondité à partir des premières naissances et renseigne sur le début de la période de procréation. Si ce sujet fait partie du recensement, il convient d'obtenir des informations pour chaque femme qui a eu au moins un enfant né vivant (Nations Unies, 2008, par. 2.193)

440. L'âge à la première naissance est déterminé soit directement, au moyen d'une donnée explicite intitulée « âge à la première naissance », ou au moyen de la différence entre l'âge du moment de la mère et l'âge de l'aîné. L'âge le plus précoce défini par chaque pays pour la procréation n'est pas l'âge biologique le plus précoce. Si un pays fixe par exemple à 13 ans l'âge minimal acceptable à la première naissance, les personnes interrogées peuvent déclarer, ou les agents de recensement peuvent noter un âge à la première naissance de 11 ou 12 ans pour telle ou telle personne. Il incombe alors aux équipes de vérification de décider si elles doivent modifier l'âge acceptable minimal, supprimer la naissance déclarée ou modifier l'âge de la mère ou encore son âge à la première naissance (sur la base, soit de l'âge d'un enfant, soit de son âge, en fonction des variables employées pour déterminer la différence d'âge). De la même manière, il incombe aux équipes de vérification de contrôler l'âge maximal à la première naissance. Vu que les femmes peuvent procréer jusqu'à 50 ans et plus, la question ne se pose pas très souvent, de telle sorte que pour corriger les erreurs éventuelles, les vérificateurs doivent déterminer la véracité des valeurs aberrantes relevées.

441. Il ne faut pas oublier que les âges minimaux et maximaux à la première naissance (ainsi que la différence d'âge entre la mère et l'enfant le plus âgé vivant au foyer familial) doivent être conformes aux coutumes et aux traditions nationales. Les spécialistes doivent donc décider lorsqu'une valeur s'apparente à un bruit et non à une valeur fondée de l'âge à la première naissance. Une fois les règles à suivre établies, les spécialistes doivent décider de la façon de remédier aux problèmes rencontrés. Si l'on ne procède pas par imputation dynamique, le programme doit attribuer la valeur « inconnu ». En revanche, si l'on utilise la méthode d'imputation dynamique, celle-ci permet de déterminer l'âge à la première naissance d'après les données d'autres femmes d'âge voisin et qui ont eu sensiblement le même nombre d'enfants. Par ailleurs, les spécialistes qui établissent la matrice d'imputation souhaitent parfois tenir compte de facteurs tels que la résidence en zone urbaine/rurale (lorsque la fécondité diffère d'une

zone à l'autre), la participation des femmes à la population active (bien que la situation du moment à cet égard ne soit pas nécessairement la même qu'au moment de la première naissance) et enfin, le niveau d'instruction.

## 9. Mortalité

442. Les renseignements sur les décès survenus au cours des 12 derniers mois servent à estimer le niveau et la structure de la mortalité par sexe et par âge dans les pays où l'enregistrement des faits d'état civil ne livre pas de statistiques suivies et de bonne qualité sur la mortalité. Pour tirer des estimations fiables de cette rubrique, il faut absolument obtenir les déclarations les plus complètes et les plus exactes possibles sur les décès par sexe et par âge au cours des douze derniers mois. C'est parce que les questionnaires des recensements réalisés ces dernières décennies comportaient des questions détaillées sur la mortalité qu'on a vu s'améliorer les résultats des méthodes d'estimation indirecte pour le calcul de la mortalité adulte (Nations Unies, 2007, par. 2.194).

443. En théorie, il faudrait pour étudier la mortalité déterminer le nombre total de décès survenus au cours des douze mois précédant la date du recensement. Lorsqu'il n'est pas possible d'obtenir des renseignements sur les décès au cours des douze derniers mois, il serait judicieux d'en recueillir au moins sur les décès d'enfants de moins d'un an. Pour chaque personne décédée, il faudrait obtenir le nom, l'âge, le sexe et la date du décès (jour, mois et année). Il convient de bien préciser la période de référence pour éviter que les recensés commettent des erreurs d'interprétation à ce sujet. On peut, par exemple, définir une période de référence précise en se servant de la date d'une fête ou d'un événement historique particulier au pays (Nations Unies, 2007, par. 2.195).

### a) Âge et sexe de la personne décédée

444. La publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* (Nations Unies, 2007) propose de recueillir le nom, l'âge et le sexe, ainsi que le jour, le mois et l'année du décès des personnes décédées pendant l'année précédant le recensement. Les pays qui n'utilisent pas la méthode d'imputation dynamique peuvent attribuer la valeur « inconnu » à chacune des variables en cas de non-validité des données. Par contre, les pays qui utilisent l'imputation dynamique pourraient employer l'âge (définition de groupes d'âge), le sexe et l'année du décès comme dimensions des matrices d'imputation des autres variables. Les autres matrices d'imputation sont vraisemblablement propres au pays, et l'équipe de vérification devra se concerter pour établir les matrices d'imputation appropriées. La structure démographique du pays ou de ses sous-régions géographiques pourrait être utile pour l'établissement de la vérification la plus appropriée.

### b) Cause du décès

445. Certains pays recueillent à présent des informations sur la cause du décès pour les décès survenus dans les 12 mois précédant le recensement. En raison du caractère délicat de cette question, et parfois du fait qu'il est difficile de recueillir de telles informations sur le terrain, les pays peuvent poser la question en ces termes : Le décès était-il dû à un accident ou à un acte de violence ? pour obtenir indirectement des renseignements sur le VIH/sida parmi certains groupes d'âge. Si l'information n'est pas recueillie ou se révèle non valide, l'équipe de vérification devra généralement décider d'affecter à cette rubrique la mention « inconnu ». Si le pays choisit de recourir à l'im-

putation, une imputation dynamique utilisant le sexe et les groupes d'âge 0, 1 à 4 ans et 5 ans serait appropriée.

#### c) Mortalité maternelle

446. Dans la série actuelle de recensements, un nombre croissant de pays demandent également si la personne décédée était de sexe féminin et si elle était enceinte au moment du décès. Cette rubrique permet de déterminer les niveaux nationaux et régionaux de mortalité maternelle. En cas de réponses non valides ou laissées en blanc, il pourrait être nécessaire d'affecter la mention « inconnu » à cette rubrique. Toutefois, si le pays choisit de recourir à l'imputation; l'imputation dynamique ne serait bien entendu appliquée qu'aux femmes, et seulement aux femmes en âge de procréer, probablement âgées de 12 à 54 ans, et probablement par année d'âge plutôt que par groupe d'âge de 5 ans.

#### d) Mortalité infantile

447. Enfin, la publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* (Nations Unies, 2008, par. 2.191) suggère la collecte d'informations sur les décès d'enfants nés « dans les 12 derniers mois ». En principe, cette question ne serait posée que dans le cadre d'une rubrique sur les naissances survenues dans les 12 mois ayant précédé le recensement. Si l'autre rubrique concernant la fécondité, la date de la dernière naissance, est utilisée, la rubrique précédente n'aurait probablement pas à être utilisée.

448. Les données sur les décès liés aux naissances survenues au cours de l'année précédant le recensement permettent aux pays ayant de bonnes informations sur les naissances et décès de vérifier leurs taux de mortalité infantile, tant au niveau national qu'au niveau régional. Là encore, une vérification des naissances d'enfants décédés au cours de l'année précédant le recensement par rapport aux décès d'enfants de moins d'un an fournirait des informations utiles sur la mortalité infantile.

449. Les vérifications de cette rubrique nécessitent une certaine réflexion et tendent à ne s'appliquer qu'au pays concerné. Théoriquement, les informations sur les enfants nés et vivants peuvent être utilisées pour vérifier les informations recueillies; lorsqu'il n'y a qu'une femme adulte dans le ménage, cette vérification est relativement simple à effectuer. Lorsque plusieurs femmes vivent dans le logement, il faut veiller à ce que le lien entre le nombre d'enfants et leur mère soit correct.

### 10. Orphelins de mère ou de père (P5G) et numéro de ligne de la mère

450. Pour recueillir des renseignements sur les orphelins, il faut, par deux questions directes, établir : *a*) si la mère naturelle de la personne recensée dans le ménage est encore vivante au moment du recensement; et *b*) si le père naturel de la personne recensée dans le ménage est encore vivant au moment du recensement. Le but est d'obtenir des informations sur les parents biologiques et donc de veiller à ne pas inclure les parents adoptifs et les parents nourriciers. Du fait qu'il existe en général plus d'un enfant survivant au sein d'une fratrie qui peut indiquer qu'il est orphelin, il convient de donner aux questions une forme telle qu'on évite ensuite de compter autant de père et de mères que d'enfants. On pose pour cela deux questions supplémentaires : *c*) le recensé est-il l'enfant survivant le plus âgé de sa mère; et *d*) le recensé est-il l'enfant survivant le plus âgé de son père (Nations Unies, 2008, par. 2.199).



451. Les vérifications concernant « la survie de la mère » et le « numéro de ligne de la mère » sont interdépendantes et doivent être réalisées conjointement. Pour les personnes répondant autrement par oui à la question de savoir si leur mère est vivante, il faut vérifier la validité du numéro de ligne de la mère; si cette donnée est valide, le code correspondant à la réponse « oui » doit être attribué à la rubrique « survie de la mère »; pour ces mêmes personnes, il faut également vérifier si le numéro de ligne de la mère est égal à 00 ou au numéro de ligne d'une femme d'au moins 12 ans. Dans l'une ou l'autre de ces deux situations, le programme suppose que la personne a une mère et attribue la réponse « oui » à la question concernant la survie de la mère. Par contre, si le numéro de ligne de la mère n'est pas valide et si la réponse à la question concernant la survie de la mère est « non » ou « ne sait pas », il faut alors supprimer le numéro de ligne indiqué. Dans tous les autres cas, il convient d'attribuer le code correspondant à la réponse « ne sait pas » à la rubrique « survie de la mère » et de supprimer toute valeur indiquée pour le numéro de ligne.

452. Le pays peut décider de ne pas vérifier le numéro de ligne de la mère relatif aux personnes dont la réponse est « non » ou « ne sait pas » à la question de la survie de la mère. Dans tous les autres cas, le numéro de ligne pourrait être vérifié quant à sa cohérence ou devrait être imputé au moyen du lien avec le chef de ménage de la personne considérée, ainsi que du numéro de ligne, du sexe, du lien avec le chef de ménage et de l'âge de la personne indiquée comme étant la mère. S'il y a des incohérences, ou si la mère ne peut être identifiée, le code correspondant à la valeur « vivant ailleurs » pourrait être attribué. Il convient de noter que, dans les vérifications structurelles, si le chef de ménage n'est pas la première personne et est alors transféré au rang de première personne, le numéro de ligne de la mère peut avoir à être ajusté pour une personne ou plus.

## B. Caractéristiques des migrations

453. Les caractéristiques démographiques d'un pays évoluent dans le temps du fait de l'accroissement naturel (fécondité et mortalité) et du solde migratoire. Les migrations peuvent être à long terme (depuis la naissance) ou à court terme, déterminées soit en fonction du précédent lieu de résidence et de la durée de séjour, soit pour une époque antérieure bien définie. Ces informations étant souvent interdépendantes, il y a lieu parfois, pour certains pays, d'effectuer une vérification conjointe semblable à celle décrite pour les variables démographiques de base. Suivant l'approche descendante, l'ordre d'exécution des vérifications joue un rôle important car certains éléments doivent être vérifiés avant d'autres.

454. Les informations sur les migrations exigent souvent des codes plus détaillés que d'autres éléments car, pour les besoins stratégiques et pratiques des pouvoirs publics, l'utilisation de subdivisions géographiques plus petites est parfois nécessaire. Ainsi, la planification des effectifs pour une nouvelle école, ou un nouveau centre de soins peut exiger des informations détaillées sur des zones peu étendues. De plus, le recours à des vérifications et à des systèmes de codage différents peut s'avérer indispensable pour les lieux situés à l'intérieur et à l'extérieur du pays.

455. Par le passé, la plupart des pays ne connaissaient pas une forte immigration internationale, de sorte qu'ils portaient leur attention sur les migrations internes, qui restent encore leur principale préoccupation. Toutefois, dans un contexte de plus en plus mondialisé, les migrations internationales prennent de plus en plus d'importance.

456. Pour les migrations internes, il faut vérifier la cohérence des données sur le lieu de naissance dans le pays et sur le nombre d'années passées dans la circonscription administrative, car il existe des liens évidents entre ces deux rubriques. De plus, il y a un certain lien logique entre les réponses fournies par les différents membres du ménage. Par exemple, si aucune réponse n'est mentionnée quant au nombre d'années passées dans la circonscription administrative par un enfant, il est possible d'imputer cette information à partir de la réponse obtenue pour la mère, puis de vérifier que la valeur imputée ne dépasse pas l'âge de l'enfant.

457. Pour les migrations internationales, il faut indiquer le pays de naissance et l'année d'entrée dans le pays d'immigration.

## 1. Lieu de naissance

458. Le lieu de naissance est d'abord le pays où la personne est née. Il n'existe pas forcément de rapport entre le pays de naissance et la nationalité, qui constitue un sujet distinct (voir Nations Unies, 2008, par. 2.92 à 2.102 et la section du présent manuel consacrée au pays de naissance). Pour les personnes nées dans le pays où est effectué le recensement (autochtones), la notion de lieu de naissance désigne également la subdivision géographique précise où résidait la mère au moment de la naissance. Dans certains pays, toutefois, le lieu de naissance pour les autochtones désigne la subdivision géographique où la naissance a effectivement eu lieu. Chaque pays aura à spécifier la définition qu'il aura choisie pour le recensement (Nations Unies, 2008, par. 2.57).

### a) Liens entre les données concernant le lieu de naissance et le nombre d'années passées dans la circonscription administrative

459. En raison des rapports étroits qui existent entre ces deux rubriques, la cohérence des données concernant le lieu de naissance et la durée de séjour peut être vérifiée. De plus, les informations concernant les différents membres d'un ménage sont liées entre elles, et des hypothèses peuvent être formulées d'après les renseignements sur les autres membres de la famille sur le point de savoir si la personne considérée a ou non émigré à l'étranger.

### b) Attribution de la valeur « inconnu » aux données non valides concernant le lieu de naissance

460. Si un pays décide de ne pas utiliser la méthode d'imputation dynamique, la valeur « inconnu » doit être attribuée à toute réponse non valide concernant le lieu de naissance. D'ordinaire, sauf omission du code, il n'y a pas lieu de procéder à une vérification de la cohérence des réponses incohérentes au sein d'une même famille ou à une vérification des zones géographiques mentionnées.

### c) Utilisation de la méthode d'imputation statique pour le lieu de naissance

461. L'indication du pays de naissance doit être modifiée uniquement lorsqu'elle est aberrante. Si la rubrique « années de séjour dans la localité » est codée par la mention « toujours », le code du pays doit être attribué au pays de naissance. Si cette rubrique contient une indication différente de « toujours », il est possible d'utiliser l'information relative à une personne précédente. Par exemple, si la personne précédente prise en compte est la mère, le nombre d'années passées par la mère dans la circonscription administrative peut être comparé à l'âge de la personne. Si ce nombre d'années est supérieur ou égal à l'âge de la personne, le code correspondant à « ce pays »

doit être attribué; sinon, il y a lieu d'attribuer le « pays de naissance de la mère ». S'il est impossible d'attribuer le pays de naissance d'après les données concernant la mère, les données concernant les autres personnes apparentées peuvent être utilisées de manière analogue. S'il est impossible d'attribuer une valeur après ces différentes vérifications, la valeur « inconnu » peut être attribuée à la rubrique « pays de naissance ».

462. Comme les pays fournissent à présent des échantillons de leurs données au public, il importe d'attribuer un code spécifique lors de la vérification des informations laissées en blanc parce qu'elles ont été omises par les agents recenseurs qui pratiquent des sauts de champ. Autrement dit, souvent, le questionnaire dit à l'agent recenseur de sauter la question sur le lieu de naissance si la personne a toujours vécu à cet endroit. Lors de la vérification, le code applicable à cet endroit particulier devrait être attribué pour aider les utilisateurs, afin qu'ils n'aient pas, par la suite, à chercher à deux endroits lors de l'établissement de tabulations croisées.

d) Utilisation de la méthode d'imputation dynamique pour le lieu de naissance

463. Comme on l'a vu précédemment, l'indication du pays de naissance doit être modifiée uniquement si elle est aberrante. Si la mention « toujours » correspond au nombre d'années passées dans la circonscription administrative, il y a lieu d'attribuer le code « ce pays » au pays de naissance. Lorsqu'une autre mention figure, les informations concernant les autres personnes du ménage doivent alors être examinées comme source d'indices pour déterminer le pays de naissance de la personne considérée.

e) Attribution du lieu de naissance lorsque la mère d'une personne est présente

464. Si le pays de naissance est laissé en blanc ou n'est pas valide, et si la durée de résidence ne porte pas la mention « toujours », il est possible de vérifier la présence éventuelle de la mère de la personne considérée. Si celle-ci fait partie du ménage, la durée de résidence de la mère est alors examinée. Si la mention « toujours » a été indiquée pour le nombre d'années passées dans la circonscription, la mention « ce pays » doit être attribuée à la variable « pays de naissance ». Si la mère de la personne considérée n'a pas toujours vécu dans la circonscription administrative, mais si l'âge de la personne est inférieur ou égal à la durée de séjour de la mère dans la circonscription administrative en question, le programme peut également attribuer la valeur correspondante à la variable « pays de naissance ». Si l'âge de la personne est supérieur au nombre d'années passées par la mère dans la circonscription en question, et si le pays de naissance de la mère est valide, la personne se voit alors attribuer le même pays de naissance que sa mère.

f) Attribution du lieu de naissance dans le cas d'un enfant du chef de ménage

465. Si la mère de la personne ne figure pas parmi les membres du ménage, mais que la personne considérée est le fils ou la fille du chef de ménage, les renseignements tirés de l'enregistrement concernant ce dernier permettent d'effectuer plusieurs vérifications pour obtenir le lieu de naissance. Si la variable « nombre d'années passées dans la circonscription administrative » est mise à la valeur « toujours » pour le chef de ménage, le programme doit attribuer la valeur « ce pays » à la rubrique « pays de naissance » de l'enregistrement relatif à la personne. Si le « nombre d'années passées par le chef de ménage dans la circonscription administrative » n'est pas mis à la valeur

« toujours » mais est supérieur à l'âge de la personne, le programme doit alors attribuer la valeur « ce pays » à la rubrique « pays de naissance » de la personne considérée. Toutefois, si l'âge de cette dernière est supérieur au nombre d'années passées par le chef de ménage dans cette circonscription, le programme doit alors lui attribuer le pays de naissance du chef de ménage, si le code correspondant est correct.

**g) Attribution d'un lieu de naissance  
à un enfant autre qu'un enfant du chef de ménage**

466. Les imputations réalisées peuvent être relativement différentes selon que l'âge de la personne considérée est supérieur ou égal à une valeur donnée (âge X) fixée par l'équipe de vérification du pays. Si l'âge de la personne est inférieur à l'âge X, l'imputation du pays de naissance doit se faire d'après le premier enregistrement précédent concernant un enfant de moins de X années, par âge et par sexe.

**h) Attribution d'un lieu de naissance aux femmes adultes mariées**

467. S'il s'agit d'une personne d'au moins X années et de sexe féminin, le programme doit vérifier la présence d'un mari parmi les membres du ménage. S'il y a un mari et si le code de son pays de naissance est valide, le programme doit attribuer ce code à l'enregistrement de son épouse. Par contre, si le code du pays de naissance du mari n'est pas valide, il faut examiner la valeur indiquée pour le nombre d'années passées dans la circonscription. Si cette donnée relative au mari porte la mention « toujours », il faut attribuer la mention « ce pays » au pays de naissance de la femme. Enfin, si le nombre d'années passées par le mari dans la localité ne porte pas la mention « toujours », il faut alors imputer le pays de naissance de la femme en fonction de l'âge et du sexe.

**i) Attribution d'un lieu de naissance aux femmes adultes non mariées**

468. Une femme d'âge supérieur à une valeur minimale fixée par l'équipe de vérification et ne comptant pas de mari parmi les membres du ménage peut cependant être la mère d'un enfant vivant sous le même toit. Dans ce cas, le programme doit vérifier la présence de l'aîné de ses enfants. Si cette opération n'aboutit pas, le programme peut alors imputer le pays de naissance en fonction de l'âge et du sexe. Par contre, si le pays de naissance de l'enfant est correctement codé et si le nombre d'années passées dans la circonscription administrative déclaré par la mère est supérieur à l'âge de l'enfant, le programme doit imputer le pays de naissance en fonction de l'âge et du sexe. Enfin, si le nombre d'années passées par la mère dans la circonscription administrative est inférieur ou égal à l'âge de l'enfant, le programme doit lui attribuer le pays de naissance de l'enfant.

**j) Attribution d'un lieu de naissance à des individus de sexe masculin**

469. Pour déterminer le lieu de naissance d'un homme, le programme de vérification peut chercher à identifier son épouse ou, s'il s'agit du chef de ménage, ses enfants. Dans un premier temps, le programme cherche à identifier l'épouse de l'homme en question. Si cette opération aboutit et si l'homme a passé dans la circonscription un nombre d'années inférieur ou égal à la durée de résidence de sa femme, le pays de naissance de la femme lui est alors attribué. Par ailleurs, si le nombre d'années passées par l'homme dans la circonscription est supérieur à la durée de résidence de sa femme, son pays de naissance doit lui être imputé en fonction de l'âge et du sexe à l'aide d'une matrice d'imputation. Si l'homme en question est le chef de ménage, s'il a un fils ou

une fille vivant sous son toit et s'il a séjourné dans la circonscription un nombre d'années inférieur ou égal à l'âge de l'enfant, le programme doit lui attribuer le pays de naissance de son enfant. Enfin, si sa durée de séjour dans la circonscription est supérieure à l'âge de son enfant, le programme doit procéder à une imputation du pays de naissance en fonction de l'âge.

## 2. Nationalité

470. Il importe de recueillir des données sur la nationalité afin d'être en mesure de classer la population en : *a*) nationaux par la naissance; *b*) nationaux par naturalisation, que la nationalité ait été acquise par déclaration, option, mariage ou de toute autre manière; et *c*) étrangers. Il convient en outre de consigner le nom du pays dont les recensés sont des nationaux en tant que tels et de ne pas utiliser l'adjectif de nationalité car celui-ci peut parfois s'appliquer également à un groupe ethnique.

471. Le codage de l'information portant sur les pays dont les recensés sont des nationaux doit être suffisamment détaillé pour permettre d'identifier individuellement chacune des nationalités représentées parmi la population étrangère vivant dans le pays. En ce qui concerne le codage, il est recommandé au pays d'utiliser le système de codage numérique figurant dans les *Codes standard des pays et des zones à usage statistique* (Nations Unies, 1999). L'utilisation de codes normalisés pour la classification de la population étrangère par nationalité renforcera l'utilité desdites données et permettra au pays d'échanger des informations sur la population étrangère. Si un pays décide de regrouper les nationalités en catégories plus larges, il lui est recommandé d'adopter les classifications régionales et sous-régionales normalisées définies dans la publication susmentionnée (Nations Unies, 2008, par. 2.97).

### a) Vérification de la nationalité

472. La nationalité dépend des définitions établies par chaque pays. Dans la plupart des pays, les personnes nées sur le territoire national sont automatiquement des nationaux par la naissance. Par conséquent, un programme de vérification doit examiner le rapport entre lieu de naissance et nationalité et, le cas échéant, attribuer ensuite le code « nationaux par la naissance » aux personnes nées dans le pays en question.

### b) Relation entre groupe ethnique/race et nationalité

473. Certains pays recueillent en outre des informations concernant le « groupe ethnique » ou la « race », susceptibles de fournir des renseignements supplémentaires à utiliser pour déterminer la nationalité, notamment en cas de non-validité de la réponse recueillie. Pour de nombreux pays, l'origine ethnique et la nationalité des migrants de la première génération doivent présenter une cohérence pratiquement parfaite. Quant aux pays qui ont une longue tradition d'accueil de migrations internationales, cette caractéristique est sans doute moins intéressante mais peut néanmoins être prise en considération avec d'autres variables.

### c) Relation entre naturalisation et nationalité

474. Dans les pays où la nationalité peut être obtenue par naturalisation, les questions posées pour le recensement ne couvrent pas nécessairement les conditions requises à cet effet. Si, par exemple, une période de résidence est nécessaire, une rubrique intitulée « durée de résidence » pourrait servir à vérifier si cette condition de natu-

ralisation est remplie. Ensuite, lorsqu'une personne est née à l'étranger et si la réponse à la question sur la nationalité est non valide ou incohérente, les équipes de vérification peuvent convenir d'inscrire « naturalisé » sous la rubrique « nationalité ». Quant aux personnes dont la durée de résidence ne répond pas aux conditions de naturalisation, le code « étranger » leur serait attribué, selon la méthode d'imputation statique.

#### d) Relation entre durée de résidence et nationalité

475. La rubrique « durée de résidence » ne figure pas sur le questionnaire ou ne permet pas de déterminer de façon catégorique la nationalité, ou encore n'a pas été retenue par l'équipe de vérification en vue de son utilisation. Si la réponse fournie à la question sur la nationalité est non valide ou incohérente par rapport au lieu de naissance, il convient de lui attribuer la mention « inconnu » lorsqu'il n'a pas été convenu d'utiliser la méthode d'imputation dynamique. En revanche, les pays qui ont choisi d'utiliser cette méthode pour corriger les valeurs non valides doivent utiliser au moins deux caractéristiques (dont l'une serait probablement le lieu de naissance) pour obtenir des informations « connues » auprès de personnes comparables provenant de la même zone géographique.

### 3. Durée de résidence

476. La durée de résidence est la période de temps arrêtée au moment du recensement et exprimée en années révolues pendant laquelle une personne a résidé : a) dans la localité qui est son lieu de résidence habituel au moment du recensement; et b) dans la plus ou moins grande circonscription administrative où cette localité est située (Nations Unies, 2008, par. 2.64).

#### a) Vérification de la durée de résidence

477. Comme pour le pays de naissance, la durée de résidence joue un rôle important lors de l'établissement de statistiques sur la mobilité de la population. Dans certains cas, un sous-groupe de la population peut présenter une mobilité nettement supérieure à celle de la nation considérée dans son ensemble. Les opérations de vérification concernant cette rubrique tiennent compte du lieu de naissance de la personne considérée et des réponses concernant les autres membres du ménage. La « durée de résidence » doit être vérifiée conjointement avec le « lieu de résidence antérieure » ou encore avec le « lieu de résidence à une date déterminée du passé ».

#### b) Résidence de fait/de droit et durée

478. La vérification peut être différente selon qu'il s'agit d'un recensement de fait ou de droit. Le recensement de droit recueillant des informations à la résidence habituelle, la « durée de résidence » ne couvre pas nécessairement les mêmes informations qu'un recensement de fait, où les personnes sont dénombrées à leur lieu de résidence le soir du recensement. De plus, les codes employés et les vérifications effectuées doivent tenir compte des personnes qui ont « toujours » vécu au même endroit ou qui ne l'ont « jamais quitté ». Dans ces cas particuliers, le programme de vérification devrait sauter différents contrôles, notamment de cohérence.

#### c) Relation entre l'âge et la durée de résidence

479. La première phase de vérification devrait consister à vérifier la cohérence entre l'âge et le lieu de naissance, ainsi que la présence d'une indication valide quant

au nombre d'années passées dans la localité ou dans la circonscription administrative en question; en effet, le nombre d'années passées par une personne dans un lieu quelconque ne peut dépasser son âge. De plus, une personne née à l'étranger ne peut avoir toujours vécu dans ladite localité ou circonscription administrative. Le programme doit donc attribuer la valeur « toujours » au nombre d'années passées dans la localité ou la circonscription administrative lorsque ce nombre d'années est supérieur à l'âge et lorsque la rubrique « pays de naissance » contient la mention « ce pays ». Si le nombre d'années passées dans la localité ou la circonscription administrative est supérieur à l'âge mais la mention « ce pays » ne figure pas sous la rubrique « pays de naissance », il convient alors d'affecter l'âge de la personne au nombre d'années passées dans la localité ou dans la circonscription administrative : dans ce cas, on suppose que la personne est bien née à l'étranger et avait moins d'un an quand elle est arrivée.

d) Relation entre lieu de naissance et durée de résidence

480. Dans ce cas, les réponses aberrantes doivent donner lieu aux mêmes contrôles que pour le lieu de naissance. Il convient à cet effet d'identifier les précédentes personnes apparentées (mère, chef de ménage, mari, enfants). L'imputation des valeurs doit s'appuyer sur les informations ainsi recueillies. Toutefois, avant d'attribuer une valeur, il faut s'assurer de sa cohérence avec l'âge et le lieu de naissance de la personne dont l'enregistrement est vérifié.

e) Pour les personnes qui ont toujours vécu au même endroit

481. Si la réponse concernant le nombre d'années passées dans la localité ou la circonscription administrative est « toujours », mais si le pays de naissance déclaré n'est pas « ce pays », l'équipe de vérification peut décider d'attribuer l'âge de la personne à la durée de résidence dans la localité ou la circonscription administrative en question. Les spécialistes supposent alors que la personne, bien que née à l'étranger, avait moins d'un an lorsqu'elle est arrivée. L'étape suivante de la vérification consistera à vérifier la validité de la mention figurant au titre du nombre d'années de résidence dans la localité ou la circonscription administrative. Étant donné que la durée de résidence d'une personne ne peut dépasser son âge, l'âge sera donc attribué au nombre d'années passées dans la localité ou la circonscription administrative.

f) Durée de résidence d'une personne d'après celle de sa mère

482. Si un code valide n'a pas été affecté à la catégorie, le programme peut vérifier dans la série d'enregistrements la présence de la mère de la personne considérée parmi les membres du ménage. Si l'issue de la recherche est positive, l'enregistrement de la mère peut fournir des informations utiles pour l'attribution des valeurs manquantes. Si la mère de la personne a toujours vécu dans la localité ou dans la circonscription administrative et si elle a déclaré pour pays de naissance « ce pays » (comme cela devrait être le cas), le programme mentionnera « toujours » sous la rubrique nombre d'années passées par la personne dans la localité ou la circonscription administrative. Si le pays de naissance déclaré par la mère n'est pas « ce pays », même lorsque la mention « toujours » a été inscrite pour la rubrique « nombre d'années passées dans la localité ou la circonscription administrative », cela indique une erreur dans les catégories concernant la mère. Le programme ignorera ensuite le pays de naissance de la mère et attribuera la valeur de l'âge à la durée de résidence dans la localité ou la circonscription administrative. Si pour cette rubrique, la mère n'a pas indiqué la mention « toujours » mais un code valide et si l'âge de la personne est inférieur au nombre

d'années passées par la mère dans la localité ou ladite circonscription, le programme de vérification effectuera un retour en arrière et vérifiera le pays de naissance de la mère si le pays indiqué est « ce pays ». Le programme attribuera l'âge de cette personne au nombre d'années passées dans la localité ou la circonscription administrative. Toutefois, si l'âge d'une personne est supérieur ou égal à la durée de résidence de la mère dans la localité ou la circonscription administrative, le programme attribuera la durée de résidence en question de la mère à la durée de résidence de la personne considérée dans cette localité ou circonscription administrative.

#### g) Durée de résidence d'une personne d'après celle d'un enfant

483. Si la personne en question est un enfant (fils ou fille), le programme de vérification doit contrôler l'enregistrement du chef de ménage pour y trouver les indications susceptibles de faciliter l'attribution de valeurs aux données manquantes concernant la durée de résidence. Si le chef de ménage est né dans « ce pays » et a toujours vécu dans la localité ou la circonscription administrative considérée, le programme attribuera la valeur « toujours » à la rubrique durée de résidence de l'enfant dans la localité ou la circonscription administrative. Si le chef de ménage a toujours vécu dans ce lieu sans néanmoins être né dans « ce pays », l'âge de l'enfant sera attribué à la durée de résidence dans la localité ou la circonscription administrative. Par contre, si le chef de ménage n'a pas fourni la réponse « toujours » à la question sur le nombre d'années de résidence dans la localité ou la circonscription administrative, mais inscrit un code valide, cette dernière information est utilisable si elle est compatible avec l'âge mentionné dans l'enregistrement de l'enfant faisant l'objet d'une vérification. Si l'âge de l'enfant est égal ou supérieur à la durée de résidence du chef de ménage dans la localité ou la circonscription administrative, le programme considérera cette durée relative au chef de ménage comme la durée de résidence du fils ou de la fille passée dans cette même localité ou circonscription administrative. Si l'âge de l'enfant est inférieur à la durée de résidence du chef de ménage dans la localité ou la circonscription administrative, le programme attribuera une valeur selon le pays de naissance du chef de ménage. La valeur attribuée sera « toujours » si le chef de ménage est né dans « ce pays »; sinon, le programme attribuera au nombre d'années de résidence dans ladite localité ou circonscription administrative une valeur égale à l'âge du fils ou de la fille.

#### h) Durée de résidence d'une personne en l'absence de toute autre information disponible

484. En cas d'échec de toutes les tentatives ci-dessus pour obtenir une valeur valide, le programme peut attribuer la valeur « non indiqué » ou « inconnu » au nombre d'années passées dans la localité ou la circonscription administrative par cette personne. Si la valeur reste non valide, il convient d'inscrire « inconnu » sous cette rubrique en cas de non utilisation de la méthode d'imputation dynamique. En revanche, en procédant par imputation dynamique des valeurs non valides, les pays doivent utiliser au moins deux caractéristiques afin d'obtenir des informations « connues » à partir de personnes similaires de la zone géographique.

### 4. Lieu de résidence antérieure

485. Le lieu de résidence antérieure est la division administrative grande ou petite, ou le pays étranger, où la personne recensée habitait immédiatement avant de s'installer dans la division administrative où elle a maintenant son domicile habituel (Nations Unies, 2007, par. 2.67).



a) Vérification du lieu de résidence antérieure

486. La rubrique « lieu de résidence antérieure » doit être vérifiée en même temps que la « durée de résidence ». Si la personne est née à cet endroit (pays, localité ou circonscription administrative, selon la rubrique de recensement) et n'a jamais déménagé, il convient alors soit de laisser en blanc cette rubrique, soit de lui attribuer un code spécifique (« n'a jamais déménagé »). Toutefois, la présence de valeurs en blanc risque de créer des problèmes lors de l'exploitation des résultats; aussi l'équipe de vérification doit-elle décider de la meilleure façon de traiter ce cas particulier.

b) Lieu de résidence antérieure lorsque les frontières ont changé

487. Les frontières nationales évoluent dans le temps et il faut donc veiller à ce que les systèmes de codage établissent les correspondances appropriées. De plus, les codes doivent être conçus de manière à permettre des regroupements logiques. Par exemple, comme il est indiqué plus haut, dans un système de codage à 3 chiffres, le premier chiffre pourrait représenter le continent de résidence, le deuxième la région et le troisième le pays.

c) Cas où la personne n'a pas déménagé depuis sa naissance

488. Le traitement des données met en tableaux des résultats concernant certaines rubriques individuelles. Aussi les spécialistes doivent-ils vérifier qu'un code spécial correspondant à la situation « né dans ce lieu » est utilisé conjointement aux autres codes de localisation. De cette manière, le programme peut distinguer entre les personnes nées à un endroit et celles nées en un lieu donné mais qui ont déménagé tout en restant dans la même zone géographique.

d) Utilisation d'autres personnes faisant partie de l'unité d'habitation

489. Lorsque le « lieu de résidence antérieure » est non valide ou incohérent, il convient généralement d'effectuer des vérifications semblables à celles qui ont été faites pour la « durée de résidence ». Le programme de vérification peut à cet effet examiner la résidence antérieure de la mère si elle réside dans l'unité d'habitation. Il peut ensuite examiner la résidence antérieure du chef de ménage pour les enfants et les adultes dans les pays où les adultes ne déménagent pas fréquemment.

e) Aucune autre personne appropriée pour la vérification du lieu de résidence antérieure

490. En cas d'échec de toutes les tentatives ci-dessus pour obtenir une valeur valide, le programme peut inscrire la mention « non indiqué » ou « inconnu » pour le nombre d'années passées dans le lieu de résidence antérieure concernant cette personne. Si la valeur reste non valide, il convient d'inscrire la mention « inconnu » sous cette rubrique, en cas de non-utilisation de la méthode d'imputation dynamique. En revanche, les pays qui choisissent d'utiliser l'imputation dynamique pour les valeurs non valides doivent utiliser au moins deux caractéristiques pour obtenir des informations « connues » à partir de personnes similaires vivant dans la zone géographique considérée.

## 5. Lieu de résidence à une date déterminée du passé

491. Le lieu de résidence à une date déterminée du passé désigne la plus ou moins importante division administrative, le pays étranger où la personne recensée habitait à une date déterminée antérieure au recensement. La date de référence choisie doit être celle qui répond le mieux aux besoins nationaux. Dans la plupart des cas, on a jugé qu'elle devait remonter à 1 ou 5 ans avant le recensement. La première date de référence permet d'obtenir les statistiques du moment sur la migration pendant une année; la seconde est peut-être plus commode pour analyser les migrations internationales, mais elle convient moins bien à l'analyse des migrations internes. Il faut également se demander, en choisissant la date de référence, dans quelle mesure les personnes recensées pourront se rappeler avec exactitude quel était leur lieu de résidence habituel un an ou cinq ans avant la date du recensement. Dans les pays qui effectuent des recensements quinquennaux, la plupart des personnes feront sans peine des rapprochements entre cette date et l'époque du recensement précédent. Dans d'autres cas, il sera probablement plus facile pour les personnes recensées de se rappeler des faits remontant à un an qu'à cinq.

492. Toutefois, certains pays devront peut-être choisir un an ou cinq ans avant le recensement, ces intervalles pouvant dans les deux cas présenter des problèmes de mémoire. Les circonstances nationales peuvent faire qu'il soit nécessaire, dans certains pays, de prendre pour date de référence une date susceptible d'être associée à un événement marquant dont chacun se souvient. L'année d'arrivée dans le pays peut, elle aussi, être une donnée utile pour l'étude des migrations internationales (Nations Unies, 2008, par. 2.69).

493. « Le lieu de résidence à une date déterminée du passé » fait l'objet du même type de vérification que le lieu de résidence antérieure. Habituellement, la question posée porte soit sur la « durée de résidence », soit simplement sur le « lieu de résidence à une date déterminée du passé ». Si la personne est née là où a lieu le recensement (pays, localité ou circonscription administrative, suivant la rubrique du recensement) et n'a jamais déménagé, cette rubrique peut alors soit être laissée en blanc, soit recevoir un code spécifique correspondant à la situation « n'a jamais déménagé ». Comme il est indiqué précédemment, les mentions laissées en blanc peuvent poser des problèmes au stade de l'exploitation des données. Aussi doit-on alors appliquer des procédures identiques à celles concernant le lieu de résidence antérieure, décrites aux trois paragraphes précédents.

## 6. Année d'arrivée

494. La publication *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision*, divise les variables relatives aux migrations en migrations internes et migrations internationales. L'année d'arrivée désigne normalement l'année d'arrivée dans le pays depuis un lieu situé hors du pays. Par conséquent, l'année d'arrivée est une question généralement posée avec une question complémentaire concernant le lieu de résidence avant l'arrivée dans ce pays (Nations Unies, 2008, par. 2.103)

### a) Relation entre l'âge et l'année d'arrivée

495. La première partie de la vérification devrait porter sur la compatibilité entre l'âge et le lieu de naissance et veiller à ce qu'une information valide concernant l'année d'arrivée dans la localité ou la circonscription administrative. Le nombre d'années qu'une personne a passées depuis son arrivée dans la localité ou la circonscription

administrative ne peut être supérieur à l'âge de la personne. En outre, une personne qui est née à l'étranger ne peut avoir vécu toute sa vie dans la localité ou la circonscription administrative. Le programme doit attribuer la mention « toujours » à l'année d'arrivée dans la localité ou la circonscription administrative si le nombre d'années dans cette localité ou circonscription administrative est supérieur à l'âge et le pays de naissance est ce pays. Si le nombre d'années depuis l'arrivée dans la localité ou la circonscription administrative est plus élevé alors que le pays de naissance n'est pas ce pays, un moyen de traiter ce cas serait d'attribuer l'âge de la personne comme nombre d'années vécues dans la localité ou la circonscription administrative. Dans ce cas, on suppose que, bien que née à l'étranger, la personne est arrivée dans la localité ou la circonscription administrative alors qu'elle avait moins d'un an.

496. Pour faciliter l'utilisation d'exemples par le public, les bureaux de statistique devraient prévoir des codes pour « moins d'un an » et « toujours » pour cette rubrique. Le code « toujours » devrait normalement être affecté au lieu de résidence du moment, afin d'aider directement à l'établissement de tableaux. Le code « moins d'un an » permettrait aux utilisateurs de s'assurer qu'ils ont considéré toute la population dans leur tabulation croisée.

#### b) Relation entre le lieu de naissance et l'année d'arrivée

497. En cas d'indications aberrantes, il convient d'utiliser les mêmes vérifications que pour le lieu de naissance. On doit rechercher des personnes apparentées (mère, chef de ménage, mari, enfant). L'imputation devrait être fondée sur les renseignements recueillis. Toutefois, avant d'attribuer une valeur, il faut s'assurer de sa compatibilité avec l'âge et le lieu de naissance de la personne dont l'enregistrement est vérifié.

#### c) Cas des personnes qui ont toujours vécu au même endroit

498. Si la réponse concernant le nombre d'années vécues dans la localité ou la circonscription administrative depuis l'arrivée de cette personne est « a toujours vécu ici », mais le pays de naissance n'est pas « ce pays », l'équipe de vérification pourra décider d'utiliser l'âge de la personne pour déterminer l'année d'arrivée dans la localité ou la circonscription administrative. Les spécialistes supposeront que, bien que née à l'étranger, la personne est arrivée dans la localité ou la circonscription administrative alors qu'elle avait moins d'un an. La phase suivante de la vérification servira à vérifier la validité de l'année d'arrivée dans la localité ou la circonscription administrative. Comme la durée de résidence d'une personne dans la localité ou la circonscription administrative ne peut être supérieur à l'âge de cette personne, dans ce cas, l'âge attribué correspondra au nombre d'années vécues dans la localité ou la circonscription administrative.

#### d) Détermination de l'année d'arrivée d'une personne à partir de l'année d'arrivée de sa mère

499. Si la catégorie considérée ne comporte pas de code valide, le programme peut procéder à une vérification à partir des données concernant la mère au sein du ménage. Si l'on trouve l'enregistrement de la mère, on peut en tirer des informations utiles pour la détermination des valeurs manquantes. Si la mère de la personne a toujours vécu dans la localité ou la circonscription administrative, et si son pays de naissance est « ce pays » (comme il devrait l'être en pareil cas), le programme attribue la mention « toujours » au nombre d'années que la personne a vécues dans la localité

ou la circonscription administrative. Si le pays de naissance de la mère n'est pas « ce pays », même si la mention « toujours » figure pour le nombre d'années qu'elle a vécues dans la localité ou la circonscription administrative, cela indique la présence d'une erreur dans les informations sur la mère. Le programme ignorera donc le pays de naissance de la mère et attribuera l'âge sur la base de l'année d'arrivée dans la localité ou la circonscription administrative. Si la mention figurant pour l'année d'arrivée de la mère dans la localité ou la circonscription administrative n'est pas « toujours » mais est un code valide, et si l'âge de la personne est inférieur au nombre d'années depuis l'arrivée de la mère dans ladite localité ou circonscription administrative, l'équipe de vérification vérifiera le pays de naissance de la mère. Si le pays de naissance de la mère est « ce pays », le programme attribuera l'âge de cette personne au nombre d'années dans ladite localité ou circonscription administrative. Toutefois, si l'âge de la personne est égal ou supérieur au nombre d'années vécues dans la localité ou la circonscription administrative, le programme attribuera « l'année d'arrivée de la mère dans la localité ou la circonscription administrative » à l'année d'arrivée de cette personne dans ladite localité ou circonscription administrative

e) Détermination de l'année d'arrivée d'un enfant  
à partir de l'année d'arrivée du chef de ménage

500. Si la personne en question est un enfant (fils ou fille), le programme de vérification devrait rechercher dans l'enregistrement concernant le chef de ménage des informations qui pourraient aider à attribuer des valeurs aux données manquantes sur l'année d'arrivée. Si le chef de ménage est né dans « ce pays » et a toujours vécu dans la localité ou la circonscription administrative, le programme attribuera la mention « toujours » au nombre d'années vécues par l'enfant dans la localité ou la circonscription administrative. Si le chef de ménage a toujours vécu dans la localité ou la circonscription administrative mais n'est pas né dans « ce pays », l'âge de l'enfant sera attribué au nombre d'années vécues dans la localité ou la circonscription administrative. Lorsque la mention concernant l'année d'arrivée du chef de ménage dans la localité ou la circonscription administrative n'est pas « toujours » mais reste un code valide, cette information peut être utilisée si elle est compatible avec l'âge figurant dans l'enregistrement de l'enfant faisant l'objet de la vérification. Si l'âge de l'enfant est égal ou supérieur à celui déterminé par l'année d'arrivée du chef de ménage dans la localité ou la circonscription administrative, le programme utilisera cette année d'arrivée du chef de ménage comme année d'arrivée de l'enfant dans ladite localité ou circonscription administrative. Si l'âge de l'enfant est inférieur à celui déterminé d'après l'année d'arrivée du chef de ménage dans la localité ou la circonscription administrative, le programme attribuera une mention selon le pays de naissance du chef de ménage. Cette mention sera « toujours » si le chef de ménage est né dans « ce pays »; sinon, le programme attribuera l'âge du fils ou de la fille comme nombre d'années de résidence dans la localité ou la circonscription administrative.

f) Détermination de l'année d'arrivée d'une personne  
en l'absence de toute autre information disponible

501. En cas d'échec de toutes les tentatives ci-dessus pour obtenir une valeur valide, le programme peut attribuer la valeur « non indiqué » ou « inconnu » à l'année d'arrivée de cette personne dans la localité ou la circonscription administrative. Si la valeur reste non valide, il convient d'inscrire « inconnu » sous cette rubrique en cas de non utilisation de la méthode d'imputation dynamique. En revanche, en procédant par imputation dynamique des valeurs non valides, les pays doivent utiliser un nom-

bre approprié de caractéristiques afin d'obtenir des informations « connues » à partir de personnes similaires de la zone géographique.

## 7. Relation entre la durée de résidence et l'année d'arrivée

502. Il est important de noter que certains pays se concentrent sur les migrations internes et ajoutent une rubrique sur la durée de résidence (souvent, de résidence précédente). D'autres pays, qui se concentrent sur les migrations internationales, ajouteront une rubrique sur l'année d'arrivée (souvent avec indication de la résidence antérieure à la migration). La plupart des pays ont soit de massives migrations internes et peu de migrations internationales, soit l'inverse, un volume considérable de migrations internationales et peu de migrations internes. Certains pays, en revanche, ont les deux types de migrations, et incluront donc deux rubriques.

503. Lorsque ces deux rubriques sont incluses, les agents du bureau de statistique doivent prendre bien soin d'établir des vérifications qui ne présentent pas d'incohérences internes. Autrement dit, il leur faut considérer à la fois les variables concernant l'âge, la durée de résidence et l'année d'arrivée pour s'assurer que la somme de la durée de résidence et de la durée écoulée depuis l'arrivée n'est pas supérieure à l'âge. Par conséquent, les programmeurs devront considérer toutes ces trois variables en même temps.

504. Lorsqu'ils utilisent la méthode d'imputation dynamique, les agents du bureau de statistique peuvent avoir à utiliser un « *hot deck* » comprenant des tableaux à dimensions multiples pour tenir compte des divers âges et nombres d'années. En outre, lorsque la durée de résidence et l'année d'entrée sont comptées en années individuelles, l'imputation dynamique doit aussi utiliser des années individuelles, car la mise à jour d'un groupe de cinq ans pourrait être cause de conflit lors de l'imputation.

505. En outre, il faut également agir avec beaucoup de soin lorsque des données groupées pour la durée de résidence ou l'année d'entrée, ou les deux, sont recueillies lors de cette vérification et lorsque l'on établit et met en œuvre des imputations dynamiques. Les données groupées peuvent poser des problèmes de recoupement. Les pays peuvent décider que l'indication d'une mention « inconnu » pourrait être la meilleure façon de faire face à cette situation.

## 8. Résidence habituelle

506. En général, aux fins du recensement, on définit la « résidence habituelle » comme le lieu où la personne vit au moment du recensement et où elle demeure depuis un certain temps ou a l'intention d'y demeurer un certain temps (Nations Unies, 2008, par. 1.461 à 1.463). Il est recommandé que les pays appliquent un seuil de 12 mois lorsqu'ils considèrent le lieu de résidence habituelle, conformément à l'une des deux définitions suivantes :

- a) Le lieu où la personne a vécu continuellement pendant la majeure partie des 12 derniers mois (c'est-à-dire au moins six mois et un jour), à l'exclusion des absences temporaires pour cause de vacances ou de déplacement professionnel, ou le lieu où elle a l'intention de vivre encore au moins six mois;
- b) Le lieu où la personne a vécu continuellement pendant au moins les 12 derniers mois, à l'exclusion des absences temporaires pour cause de vacances ou de déplacement professionnel, ou le lieu où elle a l'intention de vivre encore au moins douze mois.

507. Les pays qui effectuent des recensements de fait peuvent toutefois inclure une rubrique supplémentaire concernant la « résidence habituelle » afin d'obtenir à la fois des informations de droit et de fait. Les vérifications de cette rubrique varient selon la situation particulière du pays. Pour les personnes qui n'ont jamais déménagé, la résidence habituelle sera la même que la résidence actuelle; ainsi, l'information manquante peut être ajoutée directement.

508. Toutefois, lorsque les données indiquent un mouvement, la situation se complique. Généralement, quand cette rubrique est laissée en blanc, les pays supposent que la résidence habituelle et la résidence actuelle sont la même et que l'agent recenseur et/ou la personne interrogée a omis de fournir cette information.

509. En revanche, lorsque les données indiquent, par la durée de résidence ou l'année d'arrivée, certains signes d'un changement de résidence, les agents du bureau de statistique peuvent chercher à élaborer des méthodes susceptibles de les aider à deviner la zone géographique ou le pays en question. Bien que la vérification dépende de la situation particulière du pays, une mention « inconnu » devrait probablement être utilisée en dernier recours.

510. Si l'agent recenseur a pour instructions de laisser cette question en blanc lorsque la résidence habituelle est la même que celle où a lieu le recensement, il conviendrait d'inscrire le code pour le lieu du recensement, lors de la vérification, sous la rubrique correspondant à la résidence habituelle. Une autre variable devrait indiquer que les vérificateurs ont effectué ce changement. Le fait d'avoir un jeu complet de codes facilitera l'établissement de tableaux complets de leurs données par les utilisateurs de l'échantillon à l'usage du public.

## C. Caractéristiques sociales

511. Les caractéristiques sociales varient d'un pays à l'autre, mais correspondent généralement à des données qui rendent compte de divers aspects de la situation socioculturelle qui prévaut dans le pays considéré. Les données relatives à l'éducation, notamment l'aptitude à lire et à écrire, la fréquentation scolaire et le niveau d'instruction, ainsi que le domaine d'étude et les diplômes de fin d'études, peuvent être classés d'après les catégories de la version révisée (1997) de la *Classification internationale type de l'éducation* (CITE) mise au point par l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO) [Nations Unies, 2008, par. 2.202 à 2.230].

### 1. Aptitude à lire et à écrire (alphabétisation) [P6A]

512. Les renseignements sur l'aptitude à lire et à écrire devraient être recueillis pour toutes les personnes âgées de 10 ans et plus. Toutefois, dans un certain nombre de pays, certaines personnes âgées de 10 à 14 ans peuvent être sur le point de savoir lire et écrire grâce à leur scolarité, et le taux d'alphabétisation pour ce groupe d'âge risque d'être trompeur. Par conséquent, pour une comparaison internationale de l'alphabétisation, les tableaux des données sur l'alphabétisation devront être établis pour toutes les personnes âgées de 15 ans et plus. Lorsque les pays recueillent des données sur les personnes plus jeunes, les tableaux sur l'alphabétisation devront au moins distinguer les moins de 15 ans et les personnes âgées de 15 ans et plus (Nations Unies, 2008, par. 2.202).

513. Il incombe à chaque pays de définir l'âge minimal pour l'établissement de tableaux de résultats sur l'aptitude à lire et à écrire; de même, les équipes de vérification doivent décider de l'âge minimal pour les vérifications portant sur ce sujet, compte

tenu de la nécessité éventuelle de tableaux supplémentaires pour usage interne. Lors de l'élaboration du questionnaire, les équipes de vérification doivent décider de l'âge minimal auquel les données pertinentes sont recueillies et du niveau d'instruction à partir duquel la question n'a plus besoin d'être posée. Si la personne interrogée a déjà atteint un certain niveau de scolarité, l'agent de recensement n'a donc vraisemblablement pas besoin de poser la question sur l'aptitude à lire et à écrire. Néanmoins, cette rubrique doit être remplie lors de la vérification afin d'aider les chercheurs et autres personnes utilisant les données à l'usage du public.

514. La vérification concernant l'aptitude à lire et à écrire contrôle en premier le nombre d'années d'étude accomplies. Si ce nombre implique l'aptitude à lire et à écrire, le code correspondant à la réponse « oui » doit être attribué. En effet, les personnes ayant atteint un certain niveau de scolarité doivent être considérées comme aptes à lire et à écrire. Ainsi, lorsque le code inscrit pour l'aptitude à lire et à écrire n'est pas valide, il convient de lui attribuer une valeur. Celle-ci doit être « non indiqué » ou établie au moyen d'une matrice d'imputation fondée sur des variables déterminées, telles que le nombre d'années d'étude et le sexe. Le « niveau le plus élevé » sera fonction de la définition adoptée par le pays de l'aptitude à lire et à écrire.

## 2. Fréquentation scolaire (P6B)

515. Il convient en principe de recueillir des renseignements sur la fréquentation scolaire pour toutes les personnes quel que soit leur âge. Cela concerne en particulier la population officiellement d'âge scolaire, comprise en général entre 5 et 29 ans, mais cette tranche d'âge pouvant varier d'un pays à l'autre selon la manière dont l'enseignement est organisé dans le pays. Lorsque la collecte de données est élargie au programme d'enseignement pré-scolaire et/ou à d'autres programmes d'enseignement et de formation réguliers organisés à l'intention des adultes dans des entreprises de production et de services, des organisations communautaires et d'autres institutions que les établissements d'enseignement, la fourchette des âges est à adapter en conséquence (Nations Unies, 2008, par. 2.209).

### a) Vérification de la fréquentation scolaire

516. Dans chaque pays, l'équipe de vérification doit décider des âges adaptés à la collecte de données sur la fréquentation scolaire. Comme la plupart des pays subdivisent également la formation scolaire en plusieurs niveaux, lorsque ces niveaux doivent faire l'objet d'une compilation par âge, les spécialistes doivent en outre décider des groupes d'âge qui doivent correspondre aux différents niveaux de scolarité. Il faut modifier alors les données enregistrées pour toutes les autres personnes. Au cas où le programme de vérification produit des réponses incohérentes pour la catégorie considérée, il faut alors modifier soit l'âge, soit les données de fréquentation scolaire. Généralement, l'âge est déterminé au moment où cette vérification est effectuée, de telle sorte que la modification concerne les renseignements sur la fréquentation scolaire. Les agents de recensement doivent avoir pour instruction d'omettre les renseignements sur la fréquentation scolaire pour les personnes dont l'âge dépasse un certain nombre d'années déterminé au préalable; dans certains pays, lorsque des personnes poursuivent à l'âge adulte des études secondaires ou supérieures, il peut être indiqué de ne pas fixer des limites supérieures à l'âge de fréquentation scolaire. On peut supposer que les réponses et les combinaisons de réponses sont testées avant le recensement, à l'occasion de tests préalables, de telle sorte que les décisions à ce sujet puissent être prises avant le déroulement du recensement proprement dit.

#### b) Fréquentation scolaire à temps plein ou à temps partiel

517. Certains pays peuvent souhaiter recueillir des informations sur la fréquentation scolaire à temps partiel ou à temps plein. Si cette rubrique figure sur le questionnaire, il peut s'avérer nécessaire de procéder à sa vérification dans le cadre de la vérification portant sur la fréquentation scolaire ou séparément.

#### c) Cohérence entre la fréquentation scolaire et l'activité économique

518. Il convient tout d'abord d'effectuer les vérifications de cohérence avec d'autres rubriques importantes, telles que la principale activité économique. Si la fréquentation scolaire figure parmi les indications sur la principale activité économique, et si une personne a déclaré que cette fréquentation constitue sa principale activité économique, la réponse « oui » doit être attribuée à la question « fréquentation scolaire », et la mention « étudiant » doit figurer sous la rubrique « principale activité économique ». Autrement dit, les réponses fournies à cette question doivent être cohérentes. Dans tous les autres cas, toute réponse valide doit être acceptée.

#### d) Attribution d'une valeur en cas de réponse non valide ou incohérente concernant la « fréquentation scolaire »

519. Si la fréquentation scolaire indiquée est aberrante et si le nombre d'années d'études accomplies est valide, il convient d'attribuer une valeur au moyen d'une matrice d'imputation en fonction de l'âge, du sexe et du nombre d'années d'études. Si le code du nombre d'années d'études n'est pas valide, il convient alors d'utiliser la donnée recueillie concernant l'aptitude à la lecture et à l'écriture, pour attribuer une valeur à la « fréquentation scolaire ». Si le code de l'aptitude à la lecture et à l'écriture n'est pas valide, la valeur à inscrire sous la rubrique « fréquentation scolaire » doit être attribuée uniquement en fonction de l'âge et du sexe.

520. Il est parfois nécessaire de définir des matrices d'imputation correspondant à différentes distributions de la fréquentation scolaire par sexe et par âge (parfois par année d'âge ou par groupe d'âge d'étendue restreinte).

### 3. Niveau d'instruction (nombre d'années d'étude ou niveau d'étude atteint)

#### a) Vérification du niveau d'instruction

521. La vérification du niveau d'instruction (nombres d'années d'étude ou niveau atteint) doit comporter les opérations suivantes : a) contrôle de cohérence entre une réponse valide et l'âge; b) imputation d'une réponse lorsque la réponse initiale est aberrante. Comme on l'a indiqué précédemment, dans les pays qui n'utilisent pas la méthode d'imputation dynamique, il convient d'attribuer la mention « non indiqué ». En revanche, dans les pays qui utilisent cette méthode, il faudra tenir compte du sexe et de l'année d'âge pour les individus plus jeunes et du sexe et du groupe d'âge pour les individus légèrement plus âgés. Les pays dont les données portent à la fois sur le nombre d'années d'étude et le niveau le plus élevé atteint sont parfois tenus de recourir à des matrices d'imputation multiples (Nations Unies, 2008, par. 2.215). On trouvera dans l'annexe I des suggestions concernant l'établissement d'un nouveau code pour le « niveau d'étude actuel » sur la base de la fréquentation scolaire et du niveau d'étude atteint.



#### b) Âge minimal compatible avec le niveau d'instruction

522. Les équipes de vérification de chaque pays doivent décider de l'âge minimal de fréquentation scolaire. Une fois cet âge fixé, le nombre d'années d'études accomplies ne doit pas d'ordinaire dépasser l'âge de la personne considérée, diminué d'une valeur constante (qui correspond à l'âge minimal de fréquentation scolaire). Là encore, il importe d'utiliser l'année d'âge des enfants, en raison du risque d'introduction d'erreurs lors de la mise à jour des matrices d'imputation avec des groupes d'âge très larges.

#### c) Relation entre l'âge et le niveau d'instruction

523. L'équipe de vérification doit par ailleurs décider du niveau de bruit acceptable dans le jeu de données. D'ordinaire, il est préférable de modifier quelques cas exceptionnels en cas d'incompatibilité entre âge et niveau d'instruction, plutôt que d'accepter un grand nombre de réponses effectivement incohérentes. Quant la réponse initiale est aberrante ou incohérente avec l'âge et lorsqu'on ne procède pas par imputation dynamique, la mention « non indiqué » peut donc être inscrite. Pour les pays qui utilisent la méthode d'imputation dynamique, une réponse peut être déterminée d'après l'âge (notamment l'année d'âge des personnes d'âge scolaire), le sexe et la fréquentation scolaire. Comme l'UNESCO distingue le niveau d'instruction de « l'aptitude à lire et à écrire », cette aptitude ne devrait probablement pas figurer comme valeur dans la matrice d'imputation.

### 4. Domaine d'étude et diplôme de fin d'études

524. On a besoin de renseignements sur les personnes par niveau d'instruction et domaine d'étude pour examiner dans quelle mesure l'offre de main-d'œuvre qualifiée, dotée de telle ou telle spécialisation, correspond à la demande sur le marché du travail. Ces renseignements sont tout aussi importants pour planifier et ajuster la capacité des différents niveaux, types et secteurs des établissements d'enseignement et des programmes de formation (Nations Unies, 2008, par. 2.23).

525. Pour les personnes de moins de 15 ans (ou dont l'âge est inférieur à une valeur prédéterminée), il ne doit pas y avoir d'informations sur le domaine d'étude et/ou le diplôme de fin d'études; pour celles âgées de 15 ans et plus, il doit y avoir une relation entre niveau de formation et domaine d'étude et/ou diplôme de fin d'études. Dans chaque cas, en présence de réponses non valides, les pays qui n'utilisent pas la méthode d'imputation dynamique peuvent inscrire la mention « inconnu ». Quant aux pays qui utilisent cette méthode, ils peuvent convenir de prendre en considération l'âge, le sexe, le niveau d'instruction et, éventuellement, l'activité professionnelle pour attribuer une valeur à inscrire sous la rubrique « domaine d'étude » et/ou diplôme de fin d'études.

### 5. Religion

526. Aux fins du recensement, la religion peut se définir soit : a) par la croyance ou tendance religieuse ou spirituelle que cette croyance soit ou non représentée par un groupe organisé; ou b) par l'affiliation à un groupe organisé qui adhère à des dogmes religieux ou spirituels déterminés. Des pays étudiant la religion par le recensement utiliseront la définition qui répond le mieux à leurs besoins et les indiqueront dans la publication relative au recensement (Nations Unies, 2008, par. 2.152).

a) Vérification concernant la religion

527. La religion figure parmi les variables dont les exemples présentés au chapitre II constituent une bonne illustration. À la différence des autres caractéristiques sociales, pour les rubriques sur la religion, « l'absence de réponse » est significative et peut correspondre à certaines personnes qui hésitent à déclarer leur religion. Il est possible d'obtenir une donnée valide (y compris l'absence de réponse) concernant un individu soit directement à partir d'un autre membre du ménage, soit à partir d'un autre chef de ménage présentant des caractéristiques similaires. L'équipe de vérification doit déterminer la séquence logique de vérification adoptée pour les autres variables sociales. Le chef de ménage doit être désigné et ses données devraient être vérifiées en premier, qu'il figure ou non en tête de liste de l'unité d'habitation. Lorsque l'information sur la religion du chef de ménage est non valide ou inconnue, il convient de procéder de la façon suivante :

b) Indication de religion omise pour le chef de ménage, mais présente pour une autre personne de l'unité d'habitation

528. La première étape consiste à déterminer si une religion valide est mentionnée pour une autre personne de l'unité d'habitation, puis à attribuer la première indication valide d'une religion.

c) Indication de religion omise pour le chef de ménage ou pour toute autre personne de l'unité d'habitation

529. Si aucune religion n'est mentionnée pour aucune personne du ménage, il faut soit attribuer la mention « inconnu » à cette rubrique (si le pays considéré n'utilise pas la méthode d'imputation dynamique), soit imputer une religion à partir du cas le plus récemment traité de chef de ménage présentant des caractéristiques similaires, notamment d'âge et de sexe, ou encore de langue, de lieu de naissance ou d'autres caractéristiques éventuellement appropriées suivant le contexte.

d) Personne autre que le chef de ménage sans indication de religion

530. Si la personne considérée n'est pas le chef de ménage et ne déclare pas de religion, l'équipe de vérification peut décider de lui attribuer la religion du chef de ménage.

## 6. Langues

531. Trois types de données sur la langue peuvent être recueillies lors d'un recensement (Nations Unies, 2008, par. 2.156), à savoir :

- La langue maternelle, définie comme la langue habituellement parlée au foyer de la personne considérée dans sa petite enfance;
- La langue habituelle, définie comme la langue parlée actuellement ou celle que l'intéressé emploie actuellement le plus souvent dans son foyer;
- L'aptitude à parler une ou plusieurs langues déterminées.

a) Vérification de la langue

532. Parmi les trois langues qui peuvent être mentionnées sur le questionnaire, les deux premières langues, maternelle et langue habituelle, sont liées. Lorsque toutes deux sont indiquées sur le questionnaire, les équipes de vérification doivent envisager de les contrôler conjointement. Si l'une d'entre elles n'est pas valide, l'autre peut servir à la présentation d'une réponse.

b) Vérification de la langue : chef de ménage

533. La langue figure parmi les variables dont les exemples présentés au chapitre II donnent une bonne illustration. Les équipes de vérification doivent définir le schéma logique de vérification employé pour les autres variables sociales, en contrôlant tout d'abord les données du chef de ménage. Si la personne dont la langue est non valide ou inconnue (langue maternelle ou langue usuelle) est le chef de ménage, il faut déterminer d'abord si une langue valide est indiquée pour une autre personne de l'unité d'habitation et attribuer ainsi la première langue valide. Si aucune langue valide n'est mentionnée, il faut alors soit attribuer la mention « inconnu » si l'imputation dynamique n'est pas utilisée, soit imputer une langue d'après les données concernant le dernier chef de ménage pris en compte doté de caractéristiques similaires, notamment l'âge et le sexe, ainsi que les autres variables appropriées selon le contexte, concernant la langue et le lieu de naissance en particulier.

c) Vérification de la langue : personnes autres que le chef de ménage

534. Si la personne n'est pas le chef de ménage et si la langue n'est pas valide, il faut alors lui attribuer la langue du chef de ménage.

d) Vérification de la langue :  
utilisation de l'origine ethnique ou du lieu de naissance

535. La langue et l'origine ethnique et, parfois, le lieu de naissance sont étroitement liés et, dans le cas de certains pays, peuvent être vérifiés conjointement. Aussi les équipes de vérification doivent-elles concevoir les codes de façon à ce qu'ils reflètent les liens entre ces variables. Selon le nombre de chiffres du code et la répartition des langues nationales et des groupes ethniques, des correspondances peuvent être établies pour faciliter l'attribution de réponses adéquates en présence de données non spécifiées ou incohérentes.

e) Vérification de la langue : langue maternelle

536. Si la langue maternelle est inconnue, mais si la personne considérée est native des Philippines, il est possible de lui attribuer une langue équivalente appropriée, le tagalog, l'ilokano ou une autre langue nationale. D'ordinaire, seul le chef de ménage fait l'objet de ce type d'attribution de la langue et le code correspondant est alors attribué aux autres membres du ménage; toutefois, l'équipe de vérification de chaque pays doit tenir compte du contexte particulier, notamment des données géographiques (par exemple, résidence urbaine ou rurale), de l'âge ou d'autres données.

f) Vérification de la langue : aptitude à parler une langue déterminée

537. L'aptitude à parler une langue déterminée est une troisième variable dont les exemples présentés au chapitre II constituent une bonne illustration. Là encore, la

vérification doit porter d'abord sur le chef de ménage. S'il est fait état pour ce dernier d'une langue non valide ou inconnue, la première étape doit être de déterminer si une autre personne de l'unité d'habitation fait l'objet d'un code valide concernant l'aptitude à parler la langue en question, puis à attribuer la première langue valide observée. Par ailleurs, si aucune personne de ce type n'est identifiée, il convient alors d'attribuer à cette variable la mention « inconnu » si le pays considéré n'utilise pas la méthode d'imputation dynamique ou sinon, à imputer l'aptitude à parler une langue déterminée en fonction des données concernant le dernier chef de ménage interrogé présentant des caractéristiques similaires (par exemple, âge et sexe, mais aussi lieu de naissance et d'autres variables, le cas échéant, compte tenu du contexte). Si la personne n'est pas le chef de ménage et si la valeur déclarée pour l'aptitude à parler une langue désignée est non valide, il faut alors attribuer l'aptitude déclarée pour le chef de ménage.

## 7. Appartenance ethnique et peuples autochtones

538. La décision, lors d'un recensement, de recueillir et de diffuser des informations sur les groupes ethniques ou nationaux d'une population est fonction d'un certain nombre de facteurs et de conditions nationales, notamment, par exemple, des besoins nationaux de recueillir de telles données et de considérations de bien-fondé et de sensibilité liées au fait de poser des questions d'appartenance ethnique dans le cadre du recensement. L'identification des caractéristiques ethnoculturelles de la population d'un pays prend une importance accrue avec les migrations, l'intégration et les politiques concernant les groupes minoritaires (Nations Unies, 2008, par. 2.160).

539. La publication *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* (par. 2.163) suggère que l'on fasse preuve d'une prudence particulière lors de l'identification de peuples autochtones, qui constituent généralement une subdivision de la rubrique « appartenance ethnique ». Il faut faire très attention, lors de l'établissement de listes de codes, à veiller à ce que le caractère autochtone soit identifié uniquement pour permettre d'établir des vérifications et des tableaux afin d'aider aux travaux de planification et de formulation de politiques concernant les peuples autochtones. Par exemple, on peut avoir à définir des codes distincts pour un même groupe s'il s'agit d'une population itinérante et non pas sédentarisée dans une zone de résidence. Des vérifications spéciales peuvent être conçues, en partie en consultant des dossiers sur certains groupes particuliers de population autochtone, afin de s'assurer qu'ils sont convenablement et pleinement identifiés pour les tableaux suivants. Des procédures d'imputation spéciales peuvent être définies pour ces groupes, ou des catégories supplémentaires peuvent être utilisées dans le cadre d'imputations dynamiques existantes.

### a) Vérification concernant l'appartenance ethnique

540. Plusieurs autres variables peuvent faciliter, du moins si les données correspondantes sont recueillies, la « détermination » du groupe ethnique lorsque celui-ci est non valide ou inconnu. Dans de nombreux pays, il existe une relation entre le lieu de naissance, à l'intérieur des frontières comme à l'étranger, et le groupe ethnique. De même, la « langue maternelle » est souvent un bon indicateur du groupe ethnique pour nombre de pays, puisque les catégories correspondantes, et par conséquent les codes, seront semblables sinon identiques.

b) Vérification concernant l'appartenance ethnique du chef de ménage

541. L'origine ethnique figure parmi les variables dont les exemples présentés au chapitre II constituent une bonne illustration. Les équipes de vérification doivent définir le schéma logique de vérification employé pour les autres variables sociales, en contrôlant tout d'abord les données du chef de ménage. Si la personne dont le groupe ethnique est non valide ou inconnu est le chef de ménage, il faut déterminer d'abord si un groupe ethnique valide est indiqué pour une autre personne de l'unité d'habitation et attribuer ainsi le premier groupe ethnique valide. Si aucun groupe ethnique valide n'est mentionné, il faut alors soit attribuer la mention « inconnu » si l'imputation dynamique n'est pas utilisée soit imputer un groupe ethnique d'après les données concernant le dernier chef de ménage pris en compte doté de caractéristiques similaires (âge et sexe, ainsi que langue, lieu de naissance et autres variables appropriées selon le contexte).

c) Vérification concernant l'appartenance ethnique des personnes autres que le chef de ménage

542. Si la personne considérée n'est pas le chef de ménage et si son origine ethnique n'est pas valide, on lui attribue alors l'origine ethnique du chef de ménage.

d) Vérification concernant l'appartenance ethnique : utilisation de la langue et du lieu de naissance

543. L'origine ethnique et la langue et parfois, le lieu de naissance, sont étroitement liés et, dans certains pays, peuvent faire l'objet d'une vérification conjointe. Aussi les équipes de vérification doivent-elles concevoir les codes de façon à ce qu'ils reflètent les liens entre ces variables. Selon le nombre de chiffres du code et la répartition des langues nationales et des groupes ethniques, des correspondances peuvent être définies pour faciliter l'attribution de réponses adéquates en présence de données non spécifiées ou incohérentes.

544. Par exemple, si l'origine ethnique est inconnue, mais si la personne parle l'une des langues des Philippines et si elle est native de ce pays, il est alors possible de lui attribuer une origine ethnique équivalente appropriée (philippine). D'ordinaire, seul le chef de ménage ferait l'objet de ce type d'attribution du groupe ethnique (tandis que ce même code serait attribué aux autres membres) mais l'équipe de vérification de chaque pays doit examiner le contexte particulier, notamment les données géographiques (résidence urbaine ou rurale), l'âge ou d'autres rubriques.

## 8. Incapacité

545. L'incapacité divise la population entre ceux qui souffrent d'incapacité et ceux qui n'en souffrent pas. Une personne frappée d'incapacité se définit comme une personne plus exposée que l'ensemble de la population au risque de se voir limitée dans l'exécution de certaines tâches ou la participation à des activités qui devraient normalement lui incomber. L'Organisation des Nations Unies recommande que l'incapacité soit évaluée au regard de quatre facultés : 1) la marche; 2) la vision; 3) les facultés auditives; et 4) la connaissance (Nations Unies, 2008, par. 2.351, 2.352 et 2.367 à 2.371).

546. La question utilisée pour identifier les personnes handicapées devra comporter une liste des grandes catégories d'incapacités, afin que l'on puisse vérifier la présence ou l'absence de chaque type d'incapacité chez la personne recensée. L'incapacité peut être vérifiée sur la base de la Classification internationale des handicaps, déficien-

ces et désavantages : 1) fonctionnement et incapacité, y compris en ce qui concerne les fonctions et structures corporelles (handicaps), les activités limitées (limitations) et les limites aux possibilités de participation (restrictions); et 2) facteurs liés au contexte, y compris les facteurs environnementaux et les facteurs personnels (Nations Unies, 2008, par. 2.354).

#### a) Questions posées dans le cadre du recensement concernant l'incapacité

547. Il est recommandé d'apporter un soin particulier à la rédaction des questions du recensement concernant l'évaluation d'une incapacité. Le choix des mots et la construction des questions influent grandement sur la précision avec laquelle sont identifiées les personnes souffrant d'incapacité. Chaque domaine devrait faire l'objet d'une question distincte<sup>8</sup>. Le langage utilisé doit être simple, clair et sans ambiguïté. Les termes négatifs sont toujours à éviter. Les questions sur l'incapacité doivent être posées à chaque membre du ménage, et les questions générales sur la présence de personnes handicapées au sein du ménage sont à éviter. Le cas échéant, on peut demander à une autre personne de répondre au nom du membre de la famille qui est handicapé. Ce qui importe, c'est qu'il soit rendu compte individuellement de chaque membre de la famille plutôt que de procéder par question générale. La présentation des réponses par catégories échelonnées peut aussi contribuer à améliorer l'information sur les incapacités (Nations Unies, 2008, par. 2.373).

#### b) Vérification concernant les incapacités

548. Lorsqu'une personne ne répond pas aux questions concernant les incapacités, il est difficile de déterminer si la rubrique considérée est laissée en blanc en raison de l'absence d'incapacité ou du fait de la réticence à répondre manifestée pour une raison quelconque par la personne interrogée. L'équipe de vérification d'un pays doit donc décider s'il faut vérifier la rubrique de la manière habituelle, soit en attribuant des codes « inconnu » (sans imputation dynamique), soit sur la base des réponses fournies par d'autres personnes (dans le cas contraire). Sinon, les spécialistes peuvent décider d'accepter uniquement les réponses spécifiant l'existence d'une incapacité et de considérer toute réponse non valide comme équivalente à la réponse « aucune incapacité ». Dans ce dernier cas, la méthode d'imputation dynamique ne serait pas utilisée.

#### c) Incapacités multiples

549. Les pays qui recueillent des informations concernant la présence d'incapacités multiples devront modifier le programme de vérification. Celui-ci devra suivre le nombre d'incapacités totales envisageables, ainsi que les répétitions et la distribution de ces incapacités. Comme on l'a vu précédemment, la plupart des pays estimeront qu'il ne convient pas d'utiliser les données provenant d'autres personnes pour attribuer les valeurs à inscrire sous cette rubrique; aussi, dans les cas de non-validité, il sera vraisemblablement nécessaire d'utiliser la mention « inconnu », voire « présence d'une incapacité non établie ».

#### d) Vérification concernant les causes de l'incapacité

550. L'équipe de vérification d'un pays doit décider s'il faut vérifier la rubrique de la manière habituelle, soit en attribuant des codes « inconnu » (sans imputation dynamique) soit sur la base des réponses fournies par d'autres personnes (dans le cas contraire). Sinon, les spécialistes peuvent décider de n'accepter que les réponses spéci-

<sup>8</sup> Lorsque plusieurs domaines sont combinés, par exemple dans une question portant à la fois sur la vision et les facultés auditives, ce genre de question peut troubler la personne interrogée, qui peut se demander s'il lui faut avoir ce double handicap pour répondre « oui » à cette question. Par ailleurs, il peut être utile, tant pour la planification interne que pour des comparaisons entre pays, d'affecter des chiffres aux personnes souffrant spécifiquement de ces limitations.

fiant l'existence d'une cause d'incapacité et de ne pas utiliser de matrice d'imputation dynamique.

## D. Caractéristiques économiques

551. Les informations sur l'état de l'activité économique sont à collecter en principe pour l'ensemble de la population, mais dans la pratique, pour chaque personne ayant atteint ou dépassé un âge minimal, qui est déterminé en fonction de la situation particulière de chaque pays. L'âge où s'achève la scolarité obligatoire ne devrait pas être automatiquement retenu comme âge minimal pour la collecte de ces données. Les pays dans lesquels, en règle générale, les enfants sont nombreux à participer aux travaux agricoles ou à d'autres types d'activité économique (extraction minière, tissage, petit commerce) retiendront un âge minimal plus bas que celui des pays où l'emploi des jeunes n'est pas courant.

552. Les tableaux de caractéristiques économiques devraient au moins faire la distinction entre les personnes âgées de moins de 15 ans et celles âgées de 15 ans et plus; les pays où la scolarité est obligatoire au-delà de 15 ans et où des enfants n'ayant pas atteint cet âge exercent une activité économique devraient s'efforcer d'obtenir des données sur les caractéristiques économiques de ces enfants en vue d'assurer la comparabilité internationale des données au moins pour les personnes âgées de 15 ans et plus. Par ailleurs, il arrive souvent que l'activité économique des hommes et des femmes ayant dépassé l'âge normal de la retraite ne soit pas prise en compte. Un effort particulier devra donc être fait pour y remédier lorsqu'on cherche à établir l'effectif de la population active. Il ne devrait pas normalement être fixé d'âge maximal car bon nombre de personnes ayant dépassé l'âge de la retraite exercent parfois une activité économique de façon régulière ou occasionnelle (Nations Unies, 2008, par. 2.241).

553. Chaque pays doit fixer un âge minimal de participation à l'activité économique. Les pays désireux de recueillir des données sur le travail des enfants doivent parfois choisir un âge minimal plus bas, mais ne doit pas oublier la présence d'un certain « bruit » lorsque des enfants ne faisant pas partie de la population active y sont recensés par erreur. Une fois l'âge minimal fixé, les rubriques concernant l'activité économique sont vérifiées en vue de la mise en tableaux des résultats pour les personnes âgées de X années ou plus; par conséquent, la vérification des données concernant les enfants de moins de X années sera indispensable, ne serait-ce que pour vérifier que toutes les réponses sont laissées en blanc. Pour faciliter l'établissement de tous les tableaux, toute réponse mentionnée pour des enfants de moins de X années devra être éliminée.

### 1. Situation concernant l'activité économique

554. La situation concernant l'activité économique est établie à partir de plusieurs variables économiques, dont certaines sont décrites ci-dessous. Ces variables se prêtent bien à la collecte de données, mais doivent parfois être reclassées dans d'autres catégories pour le traitement et l'analyse des données recueillies.

555. La « situation du moment au regard de l'activité » rend compte de la relation d'une personne à l'activité économique pendant une période de référence courte, par exemple, d'une semaine ou d'un jour. On considère que l'activité du moment convient davantage aux pays où l'activité économique est faiblement influencée par les facteurs saisonniers ou autres causant des variations au cours de l'année. On peut utiliser une période de référence d'une semaine ou d'un jour, qui pourrait être soit une

semaine récente, soit la dernière semaine civile complète, soit les 7 jours précédant le recensement (Nations Unies, 2008, par. 2.248).

556. D'après les Nations Unies (2008, par. 2.253), les personnes pourvues d'un emploi sont toutes les personnes ayant dépassé un âge déterminé qui, pendant une brève période de référence d'une semaine ou d'un jour : *a*) ont effectué un travail moyennant un salaire, en vue d'un bénéfice ou d'un gain familial en espèces ou en nature; ou *b*) étaient temporairement absentes d'un emploi où elles avaient déjà travaillé et avec lequel elles avaient des liens formels, ou exerçaient une activité non salariée : avaient une entreprise industrielle ou commerciale, une exploitation agricole ou une entreprise de services.

557. La population *actuellement non active* ou, de manière équivalente, les personnes n'appartenant pas à la *population active* comprend toutes les personnes qui ne sont ni *pouvues d'un emploi*, ni *au chômage* pendant la brève période de référence utilisée pour mesurer l'activité économique du moment, y compris les personnes d'un âge inférieur à l'âge minimal spécifié pour mesurer la population économiquement active (Nations Unies, 2008, par. 2.278 à 2.279).

#### a) Catégories liées à la situation concernant l'activité économique

##### i) *Population au chômage*

558. Les chômeurs comprennent, d'après la définition des Nations Unies (2008, par. 2.271), toutes les personnes ayant dépassé un âge spécifié qui, au cours de la période de référence, étaient :

- a) Sans travail, c'est-à-dire non pourvues d'un emploi salarié ou non salarié;
- b) Disponibles pour travailler dans un emploi salarié ou non salarié durant la période de référence;
- c) la recherche d'un travail : c'est-à-dire qui avaient pris des dispositions précises au cours d'une période récente spécifiée pour chercher un emploi salarié ou non salarié. Ces dispositions peuvent inclure : l'inscription à un bureau de placement public ou privé; la candidature auprès d'employeurs; des démarches sur des lieux de travail, dans des fermes ou à la porte d'usines; sur les marchés ou autres endroits publics; l'insertion ou la réponse à des annonces dans les journaux, des recherches par relations personnelles; la recherche de terrains, d'immeubles, de machines ou d'équipements pour créer une entreprise personnelle; des démarches pour obtenir des ressources financières, des permis et licences. Il serait utile dans la classification des chômeurs d'établir une distinction entre les personnes à la recherche d'un premier emploi et les autres demandeurs d'emploi.

559. En général, pour être classée dans la catégorie des chômeurs, une personne doit réunir les trois critères ci-dessus. Toutefois, lorsque les moyens classiques de recherche d'emploi présentent peu d'intérêt, que le marché du travail est très peu organisé ou étroit, que le taux d'absorption de la main-d'œuvre est insuffisant au moment considéré ou que la population active est composée en grande partie de travailleurs indépendants (pouvus d'un emploi non salarié), on peut appliquer la définition type du chômage en assouplissant le critère de recherche d'emploi. Cela concerne les pays les moins avancés, où ce critère ne permet pas de rendre compte entièrement de l'ampleur du chômage. Cet assouplissement qui, dans les cas extrêmes, peut amener à éliminer purement et simplement ce critère, laisse comme critères fondamentaux ceux de « sans travail » et de « disponibles » pour travailler (Nations Unies, 2008, par. 2.272).



560. Les vérifications concernant les indications relatives au chômage — « licencié », « à la recherche d'un emploi », « aptitude éventuelle à prendre un emploi » et « année du dernier emploi » (le cas échéant) — devraient être effectuées conjointement. De plus, les données correspondantes doivent être compatibles avec les réponses concernant l'activité économique et, dans la plupart des cas, ne devraient pas être fournies lorsque les rubriques concernant la durée de travail, le secteur d'activité, la profession, la catégorie de travailleurs sont remplies. Si les spécialistes du sujet déterminent qu'une valeur doit figurer sous la rubrique « licencié » lorsque la réponse est soit laissée en blanc, soit non valide, alors une matrice d'imputation d'après l'âge et le sexe et, éventuellement, le niveau d'instruction de la personne pourrait être utilisée.

ii) *Population à la recherche d'un emploi*

561. La vérification concernant la rubrique « à la recherche d'un emploi » devrait se faire parallèlement à celle concernant les indications « licencié » et « cause de la non-recherche d'un emploi ». Les spécialistes du sujet devraient concevoir des vérifications d'après les réponses fournies à ces questions, afin d'imputer les autres données. La vérification devrait tenir compte de la situation locale et régionale ainsi que des différentes variables recueillies par le recensement ou l'enquête.

iii) *Population actuellement inactive*

562. La population « actuellement inactive » ou « n'appartenant pas à la main-d'œuvre », comprend toutes les personnes qui n'étaient ni pourvues d'un emploi, ni au chômage pendant la brève période de référence utilisée pour déterminer l'activité du moment. Ces personnes peuvent être classées dans les catégories suivantes, selon la raison pour laquelle elles sont « actuellement inactives » :

- a) Fréquentant un établissement d'enseignement;
- b) Effectuant des tâches ménagères;
- c) Retraitées (percevant une pension ou vivant de leurs rentes);
- d) Autres.

Les vérifications concernant les personnes « actuellement inactives » ont été intégrées aux vérifications ci-dessus, relatives à l'activité économique.

iv) *Causes de la non-recherche d'emploi*

563. Cette rubrique doit être vérifiée uniquement pour les personnes enregistrées comme n'étant pas à la recherche d'un emploi; pour toutes les autres aucune mention ne doit y figurer. Sinon, lorsqu'une mention valide figure sous « profession, secteur d'activité et situation vis-à-vis de l'emploi », il faut inscrire le code associé à la situation « pourvu d'un emploi mais ne travaillant pas actuellement ». Ce code correspond à des personnes économiquement actives, qui ont un emploi mais ne travaillaient pas pendant la période de référence. Dans tous les autres cas, et en l'absence d'imputation dynamique, la mention « inconnu » peut être attribuée. Les pays qui utilisent l'imputation dynamique peuvent par contre attribuer une valeur en fonction de l'âge, du sexe et de l'activité principale.

b) *Vérification de la situation au regard de l'activité économique*

564. D'une façon générale, l'activité économique se définit selon les catégories suivantes :

Économiquement actif

- 1) Pourvu d'un emploi

- 2) Chômeur  
Économiquement inactif
- 3) Étudiant
- 4) Personne au foyer
- 5) Retraité ou rentier
- 6) Autres

i) *Personnes pourvues d'un emploi*

565. Si l'on choisit la catégorie 1 (« Personne pourvue d'un emploi »), les valeurs des variables correspondant à la durée de travail, à la profession, au secteur d'activité, à la situation vis-à-vis de l'activité économique et au lieu de travail doivent être indiquées. Sinon, le programme de vérification doit leur être appliqué et il faut leur assigner soit la mention « inconnu », soit des valeurs obtenues par imputation statique (*cold deck*) ou dynamique (*hot deck*). Si l'on choisit la catégorie 1, les rubriques « licencié », « à la recherche d'un emploi » et « dernière année d'emploi » doivent être laissées en blanc. Dans le cas contraire, il faut leur retirer la mention indiquée et laisser la rubrique en BLANC.

ii) *Activité économique des personnes sans emploi*

566. Si l'on choisit la catégorie 2 (chômeur), une valeur doit être attribuée aux rubriques « licencié », « à la recherche d'un emploi » et « dernière année d'emploi ». Si elles ne sont pas affectées d'une indication valide, il convient de les vérifier et de leur affecter la mention « inconnu » ou une valeur obtenue par imputation statique (*cold deck*) ou dynamique (*hot deck*). Si l'on choisit les catégories 3 à 6, les rubriques correspondant à la durée de travail, à la profession, au secteur d'activité, à la situation vis-à-vis de l'activité économique et au lieu de travail doivent être laissées en blanc. Dans le cas contraire, il faut leur retirer la mention indiquée et laisser la rubrique en BLANC.

iii) *Activité économique des étudiants et des retraités*

567. Si l'on choisit la catégorie 3 (étudiant), il incombe aux spécialistes du sujet de décider si la rubrique concernant la fréquentation scolaire doit contenir la mention « oui, fréquente l'école ». Dans le cas de la catégorie 5 (retraité), il leur faut décider si la personne en question doit être d'un certain âge pour être retraitée.

iv) *Activité économique non valide et déclaration de variables concernant l'emploi*

568. Si la valeur indiquée pour l'activité économique est non valide et si des valeurs sont indiquées pour les variables concernant l'emploi (durée de travail, profession, secteur d'activité et lieu de travail), la valeur à attribuer au code de l'activité économique de la personne interrogée doit être 1. Il faudra vraisemblablement une matrice d'imputation pour choisir la réponse appropriée.

v) *Activité économique non valide et déclaration de variables concernant le chômage*

569. En cas d'indication d'une valeur pour l'une des variables concernant le chômage (« licencié », « à la recherche d'un emploi » et « dernière année d'emploi »), il convient d'attribuer à la variable « activité économique un code allant de 2 à 6. Si la personne fréquente l'école, il conviendrait probablement de lui attribuer la valeur 3. S'il s'agit d'une personne âgée, la valeur à lui attribuer serait sans doute le 5. Sinon, les

spécialistes du sujet peuvent décider d'utiliser une matrice d'imputation pour attribuer la valeur appropriée.

- (vi) *Activité économique non valide  
et aucune déclaration concernant les variables économiques*

570. Si aucune réponse n'apparaît pour les rubriques concernant l'activité économique, les spécialistes du sujet choisiront probablement d'utiliser des matrices d'imputation pour déterminer la réponse la plus appropriée et imputer des valeurs aux autres rubriques économiques.

## 2. Temps de travail

571. Le temps de travail est le temps total consacré effectivement à la production de biens et de services durant la période de référence de l'activité économique dans le recensement et comprend à la fois les heures normales de travail et les heures supplémentaires. Si la période de référence est courte (par exemple, la semaine précédant le recensement), il est recommandé de mesurer le temps de travail en heures. Dans ce cas, on peut mesurer le temps de travail en demandant des informations distinctes pour chaque journée de travail. Si la période de référence est longue (par exemple, les douze mois précédant le recensement), on mesurera si possible le temps de travail en semaines ou en journées, voire en intervalles plus longs. Le temps de travail devrait également comprendre le temps consacré aux activités qui, tout en n'aboutissant pas directement à la production de biens ou de services, relève néanmoins des tâches et responsabilités afférentes au travail, notamment le temps consacré à la préparation, à la réparation ou à l'entretien du lieu ou des outils de travail. Dans la pratique, il inclura également les temps morts survenant durant ces activités, notamment le temps passé à attendre et les autres pauses de courte durée. Les temps de pause consacrés au repas, de même que le temps au cours duquel l'intéressé n'a pas travaillé pour cause de vacances, de jours fériés, de maladie ou de conflit de main-d'œuvre ne seront pas comptés dans le temps de travail (Nations Unies, 2008, par. 2.323).

572. Cette rubrique doit faire l'objet d'une vérification uniquement pour les personnes dont la réponse à la question sur l'activité économique porte la mention « pourvu d'un emploi, au travail », ou « pourvu d'un emploi non salarié, au travail ». Dans certains pays, le temps passé à s'occuper du foyer est également compté comme temps de travail. Les catégories préalablement définies par l'équipe de vérification doivent être acceptées. En l'absence d'imputation dynamique, les rubriques laissées en blanc, marquées d'un zéro ou de codes non numériques doivent être modifiées et affectées de la mention « inconnu »; les spécialistes du sujet peuvent souhaiter remplacer la variable attribuée à la valeur économique par « ne travaille pas » si le temps de travail déclaré est égal à zéro.

573. En cas de recours à la méthode d'imputation dynamique, les variables utilisées pour la matrice d'imputation comprennent au moins le groupe d'âge et le sexe, mais d'autres variables, telles que le niveau d'instruction, la profession ou les grandes catégories de secteurs d'activité, peuvent également être utilisées.

## 3. Profession

574. La profession désigne le genre de travail effectué par une personne pourvue d'un emploi (ou, dans le cas d'un chômeur, le genre de travail effectué précédemment), quelle que soit la branche d'activité économique ou la situation de l'intéressé dans la profession (Nations Unies, 2008, par. 2.301).

575. Cette donnée doit être vérifiée uniquement pour les personnes ayant déclaré être « pourvues d'un emploi ». Si la méthode d'imputation dynamique n'est pas utilisée, la mention « non indiqué » devrait être inscrite aux rubriques laissées en blanc, marquées d'un zéro ou comportant des réponses non valides.

576. Des codes sont généralement définis pour les secteurs d'activité, de telle sorte que des chiffres différents sont affectés pour les codes de professions importantes et secondaires. Les mentions manuscrites, pratiquement inévitables dans le cas de la profession, ajoutent encore au travail de codage.

577. En cas d'utilisation de la méthode d'imputation dynamique, les variables prises en compte dans la matrice d'imputation comprennent au moins le groupe d'âge et le sexe, mais d'autres variables peuvent également être utilisées pour le niveau d'instruction ou les principaux secteurs d'activité.

#### 4. Secteur d'activité économique

578. D'après la publication *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* (par. 2.306), « l'activité économique » se rapporte à l'activité de l'établissement où la personne pourvue d'un emploi travaillait pendant la période de référence fixée pour les données sur les caractéristiques économiques (ou dans lequel elle avait travaillé en dernier, si elle est au chômage). Pour plus de précisions sur le choix des emplois ou activités devant figurer dans la classification, voir le paragraphe 2.307 de cette publication.

579. Cette donnée doit être vérifiée uniquement pour les personnes ayant déclaré au titre de l'activité économique être « pourvues d'un emploi ». En cas de non-utilisation de la méthode d'imputation dynamique, inscrire la mention « non indiqué » à la place des « blancs », des zéros ou des réponses non valides.

580. Des codes sont généralement définis pour les secteurs d'activité, de telle sorte que des chiffres différents sont affectés pour les codes de professions importantes et secondaires. Les mentions manuscrites, pratiquement inévitables dans le cas de la profession, ajoutent encore au travail de codage.

581. En cas d'utilisation de la méthode d'imputation dynamique, les variables prises en compte dans la matrice d'imputation comprennent au moins le groupe d'âge et le sexe, mais d'autres variables peuvent également être utilisées, telles que le niveau d'instruction ou les principaux secteurs d'activité.

#### 5. Situation au regard de l'emploi

582. Par situation au regard de l'emploi, on entend la situation d'une personne économiquement active vis-à-vis de son emploi, c'est-à-dire le type de contrat d'emploi explicite ou implicite qui la lie professionnellement à d'autres personnes ou organisations. Les principaux critères utilisés pour définir les groupes de la classification sont le type de risque économique, notamment la solidité du lien entre la personne et son emploi et le type d'autorité qu'elle exerce sur des établissements et d'autres travailleurs. Il faut veiller à classer les actifs selon leur situation au regard de l'emploi en se référant aux mêmes emplois que ceux utilisés pour classer les personnes par profession, par activité économique et par secteur (Nations Unies, 2008, par. 2.310)

583. La population économiquement active peut être classée comme suit en fonction de la situation au regard de l'emploi (Nations Unies, 2008, par. 2.311) :

- a) Salariés, avec possibilité de séparer les salariés ayant un contrat stable (y compris les salariés habituels) des autres salariés;

- b) Employeurs;
- c) Travailleurs indépendants;
- d) Travailleurs employés au sein de leur famille;
- e) Membres de coopératives de production;
- f) Personnes ne pouvant être classées selon leur situation.

584. Les propriétaires/chefs d'entreprises qui sont normalement classés parmi les employés, mais que l'on pourrait préférer classer comme employeurs à des fins de description et d'analyse, doivent être classés séparément.

585. Cette rubrique doit être vérifiée uniquement pour les personnes ayant déclaré au titre de l'activité économique être « pourvues d'un emploi ». En cas de non-utilisation de la méthode d'imputation dynamique, inscrire la mention « non indiqué » sous les rubriques où figurent des « blancs », des zéros ou des réponses non valides. En cas d'utilisation de la méthode d'imputation dynamique, les variables prises en compte dans la matrice d'imputation comprennent au moins le groupe d'âge et le sexe, mais d'autres variables peuvent également être utilisées, telles que le niveau d'instruction ou les principaux secteurs d'activité.

## 6. Revenu

586. Les sujets de recensement visant les caractéristiques économiques de la population présentés dans la publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitat* portent essentiellement sur la population active telle qu'elle est définie dans les *Recommandations de l'Organisation internationale du Travail* (OIT), où la notion de production économique s'entend au sens indiqué dans le *Système de comptabilité nationale* (SCN) [Nations Unies, 2008, par. 2.331]. La population active comprend toutes les personnes des deux sexes qui fournissent ou sont à même de fournir la main-d'œuvre disponible pour la production de biens et de services économiques, tels que définis dans le SCN pendant une période de référence déterminée (Nations Unies, 2008, par. 2.237).

587. Dans ce cadre, le revenu désigne soit : a) le revenu mensuel en espèces ou en nature que chaque personne active tire de son travail; soit b) le revenu annuel total en espèces ou en nature des ménages, quelle qu'en soit la source. Il est extrêmement difficile de recueillir des informations fiables sur le revenu, en particulier celui tiré d'une activité indépendante ou de biens meubles ou immeubles dans le cadre d'une étude générale sur place, particulièrement d'un recensement de la population. On accroît encore les difficultés en incluant les revenus en nature dans la définition. La collecte de données relatives aux revenus (même lorsqu'on se limite aux revenus en espèces), dans le cadre d'un recensement de population soulève des problèmes spécifiques (charge de travail accrue, erreur de réponse, etc.). C'est pourquoi ce sujet est généralement considéré comme à étudier de préférence dans le cadre d'enquêtes par sondage. Les pays pourront cependant, selon leurs besoins, chercher à obtenir par le recensement quelques informations limitées sur le revenu en espèces. Ainsi circonscrits, les renseignements recueillis serviront non seulement certains objectifs immédiats du recensement, mais encore l'étude sur la répartition du revenu, de la consommation et de l'accumulation dans le secteur des ménages.

588. La publication *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision*, distingue deux types de revenus : le revenu individuel et le revenu des ménages. Ces deux rubriques doivent faire l'objet de vérifications similaires. Pour le revenu individuel, si l'on n'utilise pas la méthode d'im-

putation dynamique, il convient d'attribuer la mention « non indiqué » ou « inconnu » en cas de réponse non valide. Dans le cas contraire, la matrice d'imputation du revenu peut être établie à partir de variables telles que l'âge, le sexe, le niveau d'instruction, le secteur d'activité, la profession et différentes caractéristiques personnelles.

589. Le revenu du ménage est égal à la somme des revenus de ses membres et doit figurer dans l'enregistrement relatif à l'unité d'habitation. La vérification de cette variable selon la méthode d'imputation dynamique utilise toutefois sensiblement les mêmes éléments que la vérification concernant le chef de ménage (âge, sexe et niveau d'instruction), plutôt que les éléments se rapportant à chacune des personnes. On trouvera un complément d'examen des recodages du revenu des ménages et du revenu familial à l'annexe I.

## 7. Secteur institutionnel

590. Les questions d'ordre institutionnel concernant l'emploi ont trait à l'organisation juridique et aux principales fonctions, aux modes de comportement et aux objectifs de l'entreprise dont relève un emploi (Nations Unies, 2008, par. 2.335).

591. Il existe un lien entre certains secteurs d'activité et certaines professions et l'emploi institutionnel (entreprise, administration publique, institutions à but non lucratif et autres). Certains pays peuvent choisir de vérifier l'existence de telles relations entre les variables pour s'assurer que les tableaux établis ne font pas apparaître d'incohérences lors de la tabulation croisée de ces variables.

592. Pour la vérification, les pays qui n'utilisent pas la méthode d'imputation dynamique devront attribuer la mention « inconnu » à la variable secteur institutionnel si cette variable n'est pas connue. Les autres pays devront envisager d'utiliser l'âge et le sexe, et peut-être le grand secteur d'activité ou la profession de personnes de la même zone géographique présentant des caractéristiques similaires.

## 8. Emploi dans le secteur non structuré

593. Lorsque les activités du secteur non structuré jouent un rôle important dans la création d'emploi et comme source de revenu, certains pays utilisent les rubriques sur l'activité et les autres rubriques économiques pour identifier ce secteur (Nations Unies, 2008, par. 2.337).

594. La vérification concernant le secteur non structuré ne devrait pas présenter de difficultés. Si la participation au secteur non structuré est indépendante de la participation au secteur structuré, on peut attribuer la mention « inconnu » aux questions laissées en blanc ou aux réponses non valides, ou bien utiliser une imputation dynamique fondée sur l'âge et le sexe. Si la participation au secteur non structuré n'est pas indépendante de la participation au secteur structuré, on peut alors ajouter une autre variable dans la matrice d'imputation dynamique pour indiquer que la personne travaillait également dans le secteur structuré (Nations Unies, par. 2.343).

## 9. Lieu de travail

595. Le « lieu de travail » est l'endroit où les personnes pourvues d'un emploi au sens de leur situation du moment exercent leur activité et/ou celles pourvues d'un emploi au sens de leur situation actuelle exercent l'activité principale qui a été utilisée pour déterminer d'autres caractéristiques économiques telles que la profession, l'activité économique et la situation au regard de l'emploi. Si l'on peut utiliser les informations relatives au lieu de travail pour établir des profils de région en ce qui concerne la

population active ayant un emploi (par opposition aux profils démographiques établis en fonction du lieu de résidence), le premier objectif demeure néanmoins d'établir un lien entre les informations sur le lieu de travail et celles sur le lieu de résidence (Nations Unies, 2008, par. 2.346).

596. Comme le « lieu de travail » sert à l'établissement de statistiques concernant les déplacements quotidiens domicile-travail, il importe que toute modification introduite dans les informations indiquées rende compte des zones géographiques particulières considérées. Aussi, les équipes de vérification de chaque pays peuvent-elles convenir d'attribuer le cas échéant la mention « inconnu » en présence de données non valides et de n'étudier que les données « connues ».

597. La durée et la complexité des opérations de codage relatives à cette rubrique seront d'autant plus importantes si les mentions manuscrites sont admises et s'il faut les coder. Si les positions numériques du code sont hiérarchisées, par exemple, si le premier chiffre représente la province, le second la circonscription et ainsi de suite, le codage sera vraisemblablement plus efficace et plus précis.

598. Les informaticiens devront vérifier que les matrices d'imputation sont établies uniquement pour des lieux géographiques vraisemblables. Il est parfois indiqué d'effectuer une nouvelle imputation statique pour chaque circonscription administrative ou autre zone géographique afin de vérifier que les valeurs précédentes ne peuvent être choisies. Quant aux matrices d'imputation proprement dites, elles peuvent inclure l'âge, le sexe et sans doute de grandes catégories modifiées de professions ou de secteurs d'activité. Il faut alors probablement utiliser des matrices d'imputation différentes pour les lieux de travail situés sur le territoire national et à l'étranger.

599. Dans le présent chapitre, nous avons examiné les variables démographiques recommandées dans les *Principes et recommandations*. Aucun pays ne devrait utiliser toutes ces variables, et les variables choisies ainsi que leurs relations géographiques avec d'autres variables devraient être sérieusement testées en vase clos et dans les situations observées avant le recensement afin de s'assurer que les réponses obtenues sont complètes et fiables. Comme les rubriques concernant la population, à la différence de celles concernant l'habitat, font généralement l'objet de nombreuses combinaisons différentes de tabulations croisées, un contrôle minutieux est nécessaire.





## Chapitre V

# Vérifications concernant l'habitation

600. Les spécifications relatives aux vérifications des données sur l'habitation tiennent compte de la validité de chacune des informations recueillies comme de leur cohérence. La connaissance de liens particuliers entre les éléments d'information concernant un pays donné permet de concevoir des vérifications de cohérence propres à garantir une meilleure qualité des données à mettre en tableaux. Par exemple, une unité d'habitation ne peut pas avoir une toiture en ciment lorsque les murs sont en bambou. De même, les unités d'habitation doivent être équipées de canalisations d'eau à l'intérieur de la maison pour pouvoir déclarer une toilette munie de chasse d'eau, ou encore une baignoire ou une douche.

601. De même que pour les données sur la population, pour les données manquantes ou non valides, l'équipe de vérification doit décider soit de leur attribuer la mention « non indiqué », soit de procéder par imputation statique (*cold deck*) pour les valeurs inconnues ou autres ou encore par imputation dynamique (*hot deck*) d'après les caractéristiques d'autres unités d'habitation. Comme précédemment, l'imputation dynamique est souvent préférable car elle supprime le type d'imputation nécessaire au stade de la mise en tableaux, alors que les décisions quant aux valeurs inconnues ne peuvent se fonder que sur les informations contenues dans les tableaux. Les matrices d'imputation ainsi établies peuvent fournir des valeurs pour les données laissées en blanc ou non valides, ou pour résoudre les incohérences lorsqu'il n'existe aucune rubrique connexe comportant des réponses valides. Certains pays peuvent présenter une certaine variabilité des caractéristiques des habitats au niveau national, mais très peu au sein de la plupart des localités. D'autres peuvent présenter pour certaines rubriques particulières des variations très importantes d'une localité à l'autre, notamment entre les zones urbaines et rurales. Il faut donc tenir compte de cette variabilité lors de l'établissement des matrices d'imputation, et tout particulièrement pour les valeurs initiations d'imputation statique. L'équipe de vérification peut décider de préciser les conditions dans lesquelles une valeur doit être attribuée à une rubrique laissée en blanc d'après les informations concernant une unité d'habitation précédente dont les caractéristiques sont voisines.

602. Sauf lorsqu'un pays ne dispose pas d'informations sur l'habitation concernant les logements collectifs (de groupe), un (et un seul) enregistrement concernant l'habitation doit être attribué à chaque numéro de série (voir chapitre III, où sont présentées différentes procédures d'assurance de la qualité). Selon les décisions prises par l'équipe de vérification, le programme conçu à cet effet créera un enregistrement concernant l'habitation si celui-ci est absent. De même, le programme de vérification peut supprimer un ou plusieurs enregistrements en présence de répétitions ou d'enregistrements multiples.

603. En principe, chaque enregistrement concernant l'habitation doit faire l'objet d'une vérification sélective limitée à certaines rubriques appropriées. Les données vérifiées peuvent différer selon qu'il s'agit d'un habitat urbain ou rural, et en fonction

de plusieurs caractéristiques, notamment climatiques. Toutefois, dans la pratique, rares sont les pays qui disposent du temps et des compétences nécessaires pour concevoir et mettre en œuvre des tableaux de données multiples afin de remplacer les données manquantes ou incohérentes. Quant aux pays qui procèdent effectivement à une vérification sélective, ils sont encore moins nombreux.

604. Néanmoins, pour des raisons d'esthétique plus que pour des raisons techniques, surtout pour les rubriques sur le logement, à mesure que la vérification se fait plus complexe et plus détaillée, on s'efforce à présent davantage de veiller à ce que les zones géographiques sélectionnées n'aient que des réponses « appropriées ». Par exemple, si certaines zones géographiques d'un pays n'ont pas l'électricité, elles ne devraient pas non plus utiliser de climatiseurs, de réfrigérateurs électriques ou de cuisinières électriques. On peut concevoir une vérification pour répondre à de telles questions dans certaines zones géographiques, de manière à éviter que des anomalies n'apparaissent dans le jeu de données final. Ainsi, bien qu'il soit possible que des personnes riches habitant dans une telle zone achètent des générateurs à gaz pour suppléer à l'absence d'électricité, l'équipe de vérification peut décider de ne pas inclure ces cas dans le jeu de données.

605. Les renseignements recueillis sur le questionnaire dépendront par ailleurs du type de locaux d'habitation (unité d'habitation ou logement collectif) et du fait que le local est occupé ou inoccupé. Dans le cas des logements collectifs ou des collectivités, la vérification peut se limiter aux données propres aux collectivités et porter conjointement sur les données recueillies aussi bien dans les collectivités que dans les autres unités d'habitation.

606. Par définition, il n'existe généralement pas d'enregistrement relatif aux habitations pour les personnes dites sans abri. Si des enregistrements de ce type existent, lorsque le pays décide de leur associer des identificateurs, il peut soit traiter ceux-ci de la même manière que ceux relatifs aux locaux collectifs, soit exiger un type de vérification totalement différent, soit encore n'exiger aucune vérification.

607. Parfois, la présence d'une mention « non indiqué » est admissible pour une rubrique particulière. Il peut en être ainsi quand l'équipe de vérification d'un pays n'a pas de données de base adéquates permettant d'imputer des réponses pour une caractéristique donnée. La décision de conserver les mentions « non indiqué » doit être prise dans le respect de l'obligation de produire des tableaux de données appropriés, adaptés aux besoins de planification et d'action des pouvoirs publics. Lorsque les planificateurs ont besoin de certaines informations, dans la mesure où les données effectivement recueillies et les données portant la mention « non indiqué » ont des distributions identiques, l'attribution de valeurs déterminées aux données « non indiqué » ne devrait pas poser de problème. Par contre, lorsque ces mêmes données portant la mention « non indiqué » présentent une certaine distorsion, l'imputation postérieure à la compilation des données risque de poser des problèmes, en particulier dans le cas des zones peu étendues ou présentant des caractéristiques particulières. Par exemple, les personnes interrogées qui habitent dans un logement défini par le pays comme étant de qualité inférieure peuvent refuser de révéler certaines des caractéristiques de leur logement. Si l'agent de recensement s'abstient de le signaler, les planificateurs risquent de ne pas pouvoir introduire de programmes d'action corrective pour remédier à cette insuffisance de qualité.

608. Les vérifications concernant l'habitation sont généralement plus simples par comparaison à celles qui concernent les données démographiques car elles comportent des tabulations croisées beaucoup moins complexes. La plupart des pays utilisent seulement différents niveaux d'agrégation géographique pour établir les caracté-

ristiques individuelles en matière d'habitation. Comme on l'a indiqué précédemment, les pays qui décident de ne pas utiliser la méthode d'imputation dynamique doivent définir un identificateur pour la mention « inconnu », à employer en présence de réponses non valides ou incohérentes.

609. Dans les pays qui procèdent par imputation dynamique, l'équipe de vérification doit établir des matrices d'imputation simples dont les dimensions permettent de distinguer différentes caractéristiques de l'habitat. Dans la plupart des cas, la variable primaire la plus indiquée en vue de l'imputation dynamique est le « type de locaux d'habitation » couvrant aussi bien les unités d'habitation proprement dites que les locaux à usage collectif et notamment les types d'unités à l'intérieur de ces catégories.

610. Pour certains pays, la zone géographique peut constituer une dimension de ces matrices d'imputation. La variable « modalités de jouissance » peut également être utilisée à cet effet. Par exemple, si dans un pays la moitié des unités d'habitation sont occupées par des locataires et l'autre moitié par des propriétaires, la variable « modalités de jouissance » peut figurer parmi les dimensions de la matrice d'imputation. Cependant, si 5 % seulement des unités sont occupées en location, une autre caractéristique serait sans doute plus indiquée. La variable « modalités de jouissance » est intéressante à utiliser dans les matrices d'imputation, notamment dans les pays où les principaux régimes d'occupation représentent d'importants pourcentages du parc de logements. Parmi les autres caractéristiques à envisager figurent le type de murs et la présence de l'électricité.

611. Pour chaque pays, les variables utilisées comme dimensions des matrices d'imputation doivent impérativement correspondre aux variables du jeu de données; ainsi, dans le cas des données sur l'habitation, il faut veiller à ce que les données individuelles ainsi que leurs combinaisons permettent de distinguer les caractéristiques recueillies.

## A. Sujets de base et sujets supplémentaires

612. Dans le cadre des recensements de l'habitat, les unités de dénombrement sont : a) les immeubles; b) les locaux d'habitation; et c) les habitants de ces locaux. Les Nations Unies ont établi une liste de sujets de base qui présentent un certain intérêt général et une certaine valeur et permettent des comparaisons statistiques internationales détaillées. Pour la commodité des utilisateurs, les codes correspondants proposés ainsi qu'un certain nombre de sujets supplémentaires sont indiqués ci-après. Les sujets sont regroupés par type d'unités de dénombrement.

### 1. Locaux à usage d'habitation : type (sujet de base)

613. La classification présentée ci-dessous décrit un système de codes à trois chiffres conçu par les Nations Unies (2008, par. 2.412 à 2.454) pour grouper, par grandes catégories, les unités d'habitation et les logements collectifs présentant des caractéristiques structurelles similaires. La distribution des occupants (population) entre les divers groupes fournit de précieuses informations sur les logements disponibles au moment du recensement. La classification offre également une base de stratification utile pour les enquêtes par sondage. Les locaux d'habitation peuvent se diviser selon les catégories suivantes :

- 1 Unités d'habitation
  - 1.1 Logements classiques

- 1.1.1 Habitations de base
- 1.1.2 Habitations ne présentant pas tous les éléments de base
- 1.2 Autres unités d'habitation
  - 1.2.1 Unités d'habitation semi-permanentes
  - 1.2.2 Unités d'habitation mobiles
  - 1.2.3 Unités d'habitation improvisées
  - 1.2.4 Unités d'habitation situées dans des bâtiments permanents non destinés à l'habitation
  - 1.2.5 Autres locaux non destinés à l'habitation
- 2 Locaux d'habitation collectifs
  - 2.1 Hôtels, maisons meublées et autres foyers d'hébergement
  - 2.2 Institutions
    - 2.2.1 Hôpitaux
    - 2.2.2 Centres de détention (prisons, pénitenciers)
    - 2.2.3 Établissements militaires
    - 2.2.4 Institutions religieuses (monastères, couvents, etc.)
    - 2.2.5 Maisons de retraite, foyers pour personnes âgées
    - 2.2.6 Locaux d'habitation pour étudiants et autres établissements similaires
    - 2.2.7 Foyers pour le personnel (par exemple, pensions et foyers pour personnel infirmier)
    - 2.2.8 Orphelinats
    - 2.2.9 Autres
  - 2.3 Camps et foyers pour travailleurs
    - 2.3.1 Camps militaires
    - 2.3.2 Camps pour travailleurs
    - 2.3.3 Camps de réfugiés
    - 2.3.4 Camps pour personnes déplacées à l'intérieur de leur pays
    - 2.3.5 Autres
  - 2.4 Autres

614. Les équipes de vérification doivent mettre au point des programmes propres à contrôler la cohérence interne des données concernant l'ensemble des locaux collectifs à usage d'habitation et des unités d'habitation. Si le type de locaux à usage d'habitation est « inconnu » ou non valide, les équipes de vérification décident parfois de mettre au point un contrôle des autres variables recueillies pour pouvoir attribuer une valeur au type de locaux d'habitation. Sinon, en cas de non-validité de la valeur indiquée, il convient d'attribuer le code « inconnu » en l'absence d'imputation dynamique. En revanche, les bureaux nationaux de recensement/statistique qui utilisent cette méthode en cas de réponses non valides doivent s'appuyer sur deux caractéristiques au moins, telles que le type d'immeuble, les modalités de jouissance, le nombre de pièces, la surface de plancher ou le statut d'occupation pour obtenir des informations « connues » à partir d'unités d'habitation similaires de la même zone géographique.

## 2. Emplacement des locaux à usage d'habitation (sujet de base)

615. L'emplacement des locaux d'habitation est une variable géographique et est présentée au chapitre III avec les vérifications structurelles.

### 3. Locaux à usage d'habitation : état d'occupation (sujet de base)

616. La question de savoir si les locaux d'habitation que leurs occupants n'utilisent que périodiquement doivent être considérés comme occupés ou non occupés sera résolue différemment selon qu'on effectue parallèlement un recensement de la population de droit ou de fait. Dans les deux cas, il semble utile de déterminer autant que possible si les locaux d'habitation sont utilisés comme résidence principale ou secondaire. Cela est particulièrement important si la résidence secondaire a des caractéristiques très différentes de la résidence principale, comme c'est le cas par exemple quand les ménages d'agriculteurs quittent leur résidence permanente dans un village pour aller s'installer pendant certaines saisons de l'année dans des locaux rudimentaires situés sur les lieux d'une exploitation agricole (Nations Unies, 2008, par. 2.466). La classification recommandée pour l'état d'occupation est la suivante :

- 1 Occupé
- 2 Inoccupé
  - 2.1 Inoccupé de façon saisonnière
    - 2.1.1 Maison de vacances
    - 2.1.2 Logements pour ouvriers saisonniers
    - 2.1.3 Autres
  - 2.2 Inoccupé pour des raisons autres que saisonnières
    - 2.2.1 Résidences secondaires
    - 2.2.2 À louer
    - 2.2.3 À vendre
    - 2.2.4 À démolir
    - 2.2.5 Autres

617. Si l'unité d'habitation est occupée, le nombre d'occupants ainsi que le décompte des enregistrements démographiques doivent être différents de zéro. Si aucune personne n'est enregistrée, l'unité est inoccupée ou bien les personnes sont manquantes. Comme on l'a indiqué plus haut à propos des vérifications structurelles, il incombe aux spécialistes de concevoir des procédures afin de déterminer si l'unité est effectivement inoccupée. Si elle est recensée comme étant occupée, mais se trouve inoccupée pour le moment, il faut alors définir une méthode de nature à permettre de déterminer le type d'inoccupation, soit en la recensant par la mention « inconnu », soit en procédant par imputation dynamique. Si l'unité est recensée comme étant inoccupée alors qu'il est possible d'établir qu'elle est occupée pour le moment sur la base des informations disponibles concernant le nombre d'occupants ou le décompte des enregistrements démographiques, il faut alors modifier la variable « état d'occupation » et lui affecter la mention « occupé ».

618. Si la valeur enregistrée n'est pas valide, le nombre d'occupants étant égal à zéro et en l'absence d'enregistrements démographiques, la mention « inconnu, inoccupé » doit être attribuée quand on ne procède pas par imputation dynamique. Si la valeur de la variable est non valide, avec un nombre d'occupants différent de zéro ou en présence d'enregistrements démographiques, la mention « occupé » doit être attribuée. Les pays ayant opté pour l'imputation dynamique des valeurs non valides (pour imputer le type d'inoccupation) doivent s'appuyer au moins sur deux caractéristiques pour obtenir des informations « connues » d'après des unités d'habitation similaire de la même zone géographique; sinon, la mention « inconnu, inoccupé » peut être attribuée.

#### 4. Propriété : type (sujet de base)

619. Ce sujet traite du régime de propriété des locaux d'habitation proprement dits et non de celui du terrain sur lequel ils se trouvent (Nations Unies, 2008, par. 2.467). Il ne faut pas confondre le régime de propriété avec les modalités de jouissance. Les informations recueillies doivent permettre de déterminer si les locaux d'habitation appartiennent au secteur public (État, administrations locales, organismes publics) ou s'ils sont propriété privée (appartenant à un ménage, à une entreprise privée, à une coopérative, à une société immobilière ou autre). La question va parfois jusqu'à demander si les locaux d'habitation sont entièrement payés, achetés à tempérament ou hypothéqués. La classification selon le régime de propriété est la suivante :

- 1 Occupé par le propriétaire
- 2 Non occupé par le propriétaire
  - 2.1 Propriété publique
  - 2.2 Propriété privée
  - 2.3 Propriété collective
  - 2.4 Propriété d'une coopérative
  - 2.5 Autres

620. Si le régime de propriété est lié aux modalités de jouissance, il faut en tenir compte lors de la mise au point de la vérification; dans le cas contraire, le type de propriété est vraisemblablement indépendant des autres variables concernant l'habitation. Si la valeur attribuée aux « types de propriété » est non valide, en l'absence d'imputation dynamique, il convient de lui attribuer le code « inconnu ». Les pays qui choisissent d'imputer les valeurs non valides par imputation dynamique doivent s'appuyer au moins sur deux caractéristiques parmi lesquelles pourraient figurer les matériaux de construction des murs, les modalités de jouissance, le type d'unité d'habitation et le nombre de pièces pour obtenir des informations « connues » à partir d'unités d'habitation similaires de la même zone géographique.

#### 5. Nombre de pièces (sujet de base)

621. La pièce se définit comme un espace d'une unité d'habitation ou d'autres locaux d'habitation entouré de murs allant du plancher au plafond (ou éventuellement au toit), ou tout au moins jusqu'à une hauteur de 2 mètres au-dessus du sol, assez grand pour contenir un lit d'adulte (4 mètres carrés au moins). Sont compris dans cette définition les chambres à coucher, les salles à manger, les salles de séjour, les cabinets de travail, les greniers habitables, les chambres de domestiques, les cuisines, les pièces servant à l'exercice d'une profession ou d'un négoce et autres espaces séparés, utilisés pour l'habitation ou destinés à l'habitation, à condition qu'ils répondent aux critères relatifs à la surface habitable et aux murs. Les couloirs, vérandas, vestibules, etc., les salles de bains et cabinets d'aisances ne doivent pas être comptés comme pièces, même s'ils répondent à ces critères. À des fins nationales, on pourra recueillir des informations distinctes concernant des locaux de moins de 4 mètres carrés répondant par ailleurs à la définition d'une pièce, si l'on considère que leur nombre le justifie (Nations Unies, 2008, par. 2.472).

622. Comme le nombre de pièces peut être indépendant des autres variables concernant l'habitation, si la réponse indiquée est non valide, en l'absence d'imputation dynamique, il convient de lui attribuer le code « inconnu ». Les pays qui choisissent d'imputer les valeurs non valides par imputation dynamique doivent s'appuyer au moins sur deux caractéristiques parmi lesquelles pourraient figurer le type d'unité

d'habitation, les matériaux de construction des murs, les modalités de jouissance et l'état d'inoccupation pour obtenir des informations « connues » à partir d'unités d'habitation similaires de la même zone géographique.

## 6. Nombre de chambres à coucher (sujet supplémentaire)

623. Outre l'énumération du nombre de pièces, plusieurs recensements nationaux recueillent des informations sur le nombre de chambres à coucher par unité d'habitation, qui constitue l'unité de dénombrement de ce sujet. Une chambre à coucher se définit comme une pièce équipée d'un lit et utilisée pour le repos nocturne (Nations Unies, par. 2.475).

624. Parfois, les agents recenseurs indiquent un chiffre pour le nombre de chambres à coucher qui est supérieur au nombre de pièces<sup>9</sup>. Si cela se produit et si le pays n'utilise la mention « non indiqué » que pour les réponses non valides ou incohérentes, la mention « non indiqué » doit figurer pour le nombre de chambres à coucher. S'il fait usage de l'imputation dynamique, le nombre de chambres devrait être « estimé » à l'aide d'une matrice d'imputation où le nombre de pièces serait l'un des éléments. Ainsi, le nombre de chambres ne sera pas supérieur au nombre de pièces, car la valeur indiquée pour le nombre de chambres ne sera mise à jour que lorsque les valeurs indiquées pour les pièces et les chambres concordent. Le cas le plus simple serait une tabulation linéaire avec le nombre de pièces comme cases et la valeur indiquée pour les chambres inscrite dans les cases. Une matrice d'imputation plus complexe pour comprendre le nombre de personnes vivant dans l'unité d'habitation et le type de structure.

625. Autrement, si la valeur indiquée pour le nombre de chambres est non valide, la mention « inconnu » devrait être attribuée en l'absence d'imputation dynamique. Les pays qui choisissent d'utiliser l'imputation dynamique pour les valeurs non valides devraient utiliser au moins deux caractéristiques (l'une étant le nombre de pièces) pour obtenir une information « connue » auprès d'unités d'habitation similaires de la même zone géographique.

## 7. Espace plancher utilisable (sujet supplémentaire)

626. Par espace plancher, on entend l'espace plancher utile dans les unités d'habitation, c'est-à-dire l'espace plancher mesuré à l'intérieur des murs extérieurs, à l'exclusion des caves et des greniers non habitables. Dans les bâtiments à unités d'habitation multiples, les espaces communs sont à exclure. L'approche doit être différente selon qu'il s'agit d'unités d'habitation ou de locaux d'habitation collectifs (Nations Unies, 2008, par. 2.476).

627. L'espace plancher peut se rapporter au nombre de pièces et/ou au nombre de chambres à coucher; les équipes de vérification peuvent donc décider de tenir compte de cette différence lors de la conception des programmes de vérification. Parmi les autres rubriques utiles à l'imputation dynamique figurent le nombre d'occupants et le nombre d'occupants par pièce. Le plus souvent, l'espace plancher est indépendant des autres vérifications concernant les données sur l'habitation. L'unité de mesure utilisée, par exemple, le mètre carré, doit parfois être spécifiée. Si la valeur indiquée est non valide, en l'absence d'imputation dynamique, il convient d'attribuer la mention « inconnu ». Les pays qui choisissent de procéder par imputation dynamique pour les valeurs non valides doivent utiliser au moins deux caractéristiques, notamment le type d'unité d'habitation, les matériaux de construction des murs, les modalités de jouis-

<sup>9</sup> En cas d'indication du nombre de pièces et du nombre de chambres, ces deux données devraient être vérifiées conjointement, et le nombre de chambres ne devrait pas dépasser le nombre de pièces. Comme le nombre de chambres est un sujet « supplémentaire », il n'est procédé à la vérification que si ces deux nombres sont indiqués.

sance et le statut d'occupation pour obtenir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

10 Pour les variables suivantes, les unités de dénombrement sont effectivement le système d'alimentation en eau de l'unité d'habitation; les lieux d'aisances et l'assainissement; les salles de bains; les équipements servant à la préparation des repas; le mode d'éclairage et l'évacuation des déchets solides.

## 8. Système d'alimentation en eau (sujet de base<sup>10</sup>)

628. D'après les Nations Unies (2008, par. 2.479), les principaux renseignements à obtenir par le recensement concernant le système d'alimentation en eau sont de savoir si les unités d'habitation ont l'eau sous conduite, afin de déterminer si elles sont alimentées en eau par des canalisations reliées soit à un système qui dessert toute la communauté, soit à des installations particulières, telles que réservoir à pression ou pompe. À cette fin, l'unité de dénombrement est l'unité d'habitation. Il faut également indiquer s'il y a un poste d'eau à l'intérieur de l'unité ou, sinon, à moins d'une certaine distance de la porte d'entrée. La distance recommandée à cet égard est 200 mètres, étant entendu que l'accès à l'eau sous conduite en deçà de cette distance doit permettre aux occupants de l'unité d'habitation de se procurer l'eau nécessaire à leurs besoins ménagers sans avoir à fournir d'efforts excessifs. Outre l'emplacement du poste d'eau, la source d'approvisionnement présente un autre intérêt particulier. La classification recommandée des unités d'habitation selon leur système d'alimentation en eau est donc la suivante :

- 1 Poste d'eau sous conduite à l'intérieur de l'unité
  - 1.1 À partir d'un réseau d'approvisionnement collectif
  - 1.2 À partir d'une source d'approvisionnement individuelle
- 2 Poste d'eau sous conduite à l'extérieur de l'unité, mais à moins de 200 mètres
  - 2.1 À partir d'un réseau d'approvisionnement collectif
    - 2.1.1 À usage collectif
    - 2.1.2 À usage commun
  - 2.2 À partir d'une source d'approvisionnement individuelle
    - 2.2.1 À usage exclusif
    - 2.2.2 À usage commun
3. Autres

629. Par réseau d'approvisionnement collectif, on entend un réseau soumis à inspection et contrôle des pouvoirs publics. Ces réseaux sont généralement gérés par des organismes publics, mais il arrive qu'ils soient exploités par une coopérative ou une entreprise privée.

630. Les rubriques concernant les installations d'alimentation en eau — système d'approvisionnement, eau potable, lieux d'aisances et égouts, salles d'eau et eau chaude — devraient probablement faire l'objet d'une vérification conjointe. Comme ces données sont étroitement liées, lorsque l'une d'entre elles manque ou est non valide, les autres peuvent servir à l'obtention d'une valeur appropriée. Dans les zones géographiques sans eau courante, les spécialistes peuvent avoir à appliquer des vérifications particulières aux unités d'habitation. Sinon, d'autres unités de la région présentent probablement des caractéristiques similaires, et l'utilisation des rubriques correspondantes est recommandée en cas d'application de la méthode d'imputation dynamique.

631. Si la valeur attribuée au système d'alimentation en eau est non valide, il convient de lui attribuer la mention « inconnu » en l'absence d'imputation dynamique. En revanche, les pays ayant opté pour cette méthode doivent utiliser au moins deux caractéristiques. En règle générale, celles-ci pourraient être le type d'unité d'habitation, puis celle des lieux d'aisances et égouts et enfin, des salles d'eau, de manière à obtenir



des informations « connues » à partir d'unités d'habitation similaires de la même zone géographique.

## 9. Eau potable : principale source (sujet de base)

632. La vérification concernant l'eau potable devrait se faire conjointement avec celle du système d'alimentation en eau. Étant donné que l'eau en bouteilles et les autres sources non traditionnelles d'eau potable sont normalement incluses dans le questionnaire, elles doivent également figurer dans la vérification (Nations Unies, par. 2.483)

633. Si la valeur attribuée à l'eau potable est non valide, il convient de lui attribuer la mention « inconnu » en l'absence d'imputation dynamique. En revanche, les pays ayant opté pour cette méthode doivent utiliser au moins deux caractéristiques. En règle générale, celles-ci pourraient être le type d'unité d'habitation, puis celles du système d'alimentation en eau, des lieux d'aisances et égouts et des salles d'eau, de manière à obtenir des informations « connues » à partir d'unités d'habitation similaires de la même zone géographique.

## 10. Lieux d'aisances : types (sujet de base) et

### 11. Égouts (sujet de base)

634. La vérification concernant les lieux d'aisances et les égouts devrait être combinée à celle des autres variables concernant la plomberie pour produire des résultats plus homogènes. Bien que la publication *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* ait combiné ces deux variables, celles-ci ont été séparées pour le recensement de 2010. Néanmoins, ces rubriques devraient être vérifiées conjointement et, si possible, avec la même matrice d'imputation dynamique.

635. Dans certains pays, on a jugé utile de subdiviser la catégorie des cabinets d'aisances sans chasse d'eau pour distinguer certains types d'usage courant qui correspondent à un certain niveau d'hygiène. La classification recommandée par les Nations Unies (2008, par. 2.487) pour répertorier les unités d'habitation en fonction des lieux d'aisances est la suivante :

- 1 Cabinet d'aisances à l'intérieur de l'unité d'habitation
  - 1.1 Avec chasse d'eau/chasse d'eau manuelle
  - 1.2 Autre
- 2 Cabinet d'aisances à l'extérieur de l'unité d'habitation
  - 2.1 À usage exclusif
    - 2.1.1 Avec chasse d'eau/chasse d'eau manuelle
    - 2.1.2 Latrines à fosse améliorée avec ventilation
    - 2.1.3 Latrines à fosse sans ventilation avec couvercle
    - 2.1.4 Trous ou fosses à couvercle temporaire ou sans abri
    - 2.1.5 Autre
  - 2.2 Cabinet d'aisances à usage partagé
    - 2.2.1 Avec chasse d'eau/chasse d'eau manuelle
    - 2.2.2 Latrines à fosse améliorée avec ventilation
    - 2.2.3 Latrines à fosse sans ventilation avec couvercle
    - 2.2.4 Trous ou fosses à couvercle temporaire ou sans abri

### 2.2.5 Autre

- 3 Pas de cabinet d'aisances disponible
  - 3.1 Seau hygiénique (excréta évacués manuellement)
  - 3.2 Utilisation de l'environnement naturel, par exemple, buisson, rivière, cours d'eau, etc.

636. Le type de lieux d'aisances et d'égouts constitue une autre rubrique concernant l'habitation qui a un rapport avec l'approvisionnement en eau et devrait donc être vérifié conjointement avec les autres rubriques connexes. Le choix de mentions, telles que « à usage privé », « à usage commun », « à usage exclusif », etc., pourrait servir à déterminer la cohérence des différentes réponses et, si tel n'est pas le cas, les chemins de vérification à suivre pour résoudre le problème. En présence d'une ou de plusieurs variables concernant l'eau, il est possible d'établir une estimation concernant les données inconnues ou incohérentes sans recourir à l'utilisation du code « inconnu » ou à l'imputation dynamique. Toutefois, faute d'obtenir une valeur valide, la mention « inconnu » doit être attribuée en l'absence d'imputation dynamique. Par contre, les pays ayant opté pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, notamment, en règle générale, le type d'unité d'habitation, ainsi que l'alimentation en eau, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation, afin de recueillir des données connues à partir d'unités d'habitation similaires de la même zone géographique.

## 12. Salles d'eau (sujet de base)

637. D'après les Nations Unies (2008, par. 2.490), il convient de déterminer pour chaque local d'habitation s'il y a une baignoire ou une douche fixe. Pour cette rubrique, l'unité de dénombrement est également l'unité d'habitation. On peut établir en outre si les installations sont à usage exclusif des occupants du local ou non et si la salle d'eau est alimentée en eau chaude pour le bain ou seulement en eau froide. Il est néanmoins possible, dans certaines régions du monde, que la distinction proposée ci-dessus ne corresponde pas le mieux aux besoins du pays. Ainsi, il peut être important de faire la distinction entre une salle d'eau séparée dans le bâtiment et un établissement de bains publics. La classification recommandée des unités d'habitation en fonction des salles d'eau dont elles sont dotées ou non est la suivante :

- 1 Baignoire ou douche fixe intérieure
- 2 Pas de baignoire ou douche fixe intérieure
  - 2.1 Baignoire ou douche fixe disponible à l'extérieur de l'unité d'habitation
    - 2.1.1 À usage exclusif
    - 2.1.2 À usage commun
  - 2.2. Pas de baignoire ou de douche fixe disponible

638. Le type de salle d'eau devrait faire l'objet d'une vérification conjointe avec les autres rubriques connexes. Le choix de mentions telles que « privé », « à usage commun », « à usage exclusif », etc., pourrait servir à déterminer la cohérence des différentes réponses et, si tel n'est pas le cas, les chemins de vérification à suivre pour résoudre le problème. En présence d'une ou de plusieurs variables concernant l'eau, il est possible d'établir une estimation concernant les données inconnues ou incohérentes sans recourir à l'utilisation du code « inconnu » ou à l'imputation dynamique. Toutefois, si toutes les tentatives échouent et si la valeur est non valide, le code « inconnu » doit être attribué en l'absence d'imputation dynamique. Par contre, les pays ayant opté pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur

deux caractéristiques, notamment, en règle générale, le type d'unité d'habitation, ainsi que l'alimentation en eau, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

### 13. Cuisine : disponibilité (sujet de base)

639. D'après la publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* (Nations Unies, 2008, par. 2.494), la collecte de données sur l'existence d'une cuisine peut être l'occasion de recueillir des renseignements sur le type d'équipements utilisés pour la préparation des repas (fourneaux, plaques chauffantes, feu nu, etc.) ainsi que sur l'existence d'un évier et d'un endroit prévu pour stocker les aliments afin d'éviter qu'ils ne se gâtent. La classification recommandée des unités d'habitation selon l'existence d'une cuisine ou de tout autre espace réservé à la préparation des repas est la suivante :

- 1 Cuisine intérieure
  - 1.1 À usage exclusif
  - 1.2 À usage commun
- 2 Autre espace intérieur pour la préparation des repas, tel que kitchenette
  - 2.1 À usage exclusif
  - 2.2 À usage commun
- 3 Ni cuisine, ni espace réservé à la préparation des repas à l'intérieur de l'unité d'habitation
  - 3.1 Cuisine ou autre espace réservé à la préparation des repas à l'extérieur de l'unité d'habitation
    - 3.1.1 À usage exclusif
    - 3.1.2 À usage commun
  - 3.2 Ni cuisine, ni espace réservé à la préparation des repas.

640. Pour la vérification des données recueillies sur les installations destinées à la préparation des repas, des mentions telles que « privé », « à usage commun », « à usage exclusif » peuvent servir à déterminer la cohérence des différentes réponses et, à défaut, les chemins de vérification à suivre pour résoudre le problème. En présence d'une ou de plusieurs variables concernant les installations de préparation des repas, il est possible d'établir une estimation concernant les données inconnues ou incohérentes sans recourir à l'utilisation du code « inconnu » ou à l'imputation dynamique. Toutefois, si la valeur est non valide, le code « inconnu » doit être attribué en l'absence d'imputation dynamique. Par contre, les pays ayant opté pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, notamment, en règle générale, le type d'unité d'habitation, ainsi que l'alimentation en eau, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

### 14. Combustible utilisé pour la cuisson des aliments (sujet de base)

641. Devant la nécessité de suivre de près l'utilisation des ressources naturelles, un certain nombre de recensements nationaux de l'habitat prévoient des questions sur le combustible utilisé pour la cuisson des aliments. L'unité retenue pour le dénombrement est l'unité d'habitation, et l'expression « combustible utilisé pour la cuisson des

aliments » se réfère au combustible utilisé dans la plupart des cas pour la préparation des principaux repas. Si deux combustibles sont utilisés (par exemple, l'électricité et le gaz), il convient d'indiquer celui qui est utilisé le plus souvent. La classification des combustibles utilisés pour la cuisson des aliments dépend de la situation nationale et peut comprendre l'électricité, le gaz, le pétrole, le charbon, le bois et les déchets animaux. Il est également utile de recueillir des informations pour les locaux d'habitation collectifs, surtout si le nombre d'ensembles de ce type de locaux est élevé dans le pays (Nations Unies, 2008, par. 2.496).

642. La vérification des réponses concernant le type de combustible utilisé pour la cuisson des aliments devrait être effectuée conjointement avec celle des réponses concernant les installations utilisées à cet effet. L'équipe de vérification détermine la relation entre ces deux variables et met au point une vérification qui permet de contrôler leur compatibilité. Les mentions telles que « privé », « à usage commun », « à usage exclusif » et ainsi de suite seront probablement utilisées pour déterminer la cohérence des différentes réponses et, à défaut, les chemins de vérification à suivre pour résoudre le problème. En présence d'une ou de plusieurs variables concernant la préparation des repas, il est possible d'établir une estimation concernant les données inconnues ou incohérentes sans recourir à l'utilisation du code « inconnu » ou à l'imputation dynamique. Toutefois, si la valeur est non valide, le code « inconnu » doit être attribué en l'absence d'imputation dynamique. Par contre, les pays ayant opté pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, notamment, en règle générale, le type d'unité d'habitation, ainsi que l'alimentation en eau, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

## 15. Éclairage et/ou électricité : type (sujet de base)

643. Il convient de recueillir des informations sur le mode d'éclairage des locaux d'habitation (électricité, gaz, lampes à pétrole ou autres). Si les locaux sont éclairés à l'électricité, certains pays souhaiteront peut-être préciser si le courant provient du secteur, de l'utilisation d'une génératrice ou d'une autre source (usine). Outre le type d'électricité utilisé, les pays devront chercher à déterminer si l'électricité est utilisée à d'autres fins de l'éclairage (notamment pour la préparation des repas, l'eau chaude ou le chauffage). Si la situation du logement dans le pays fait qu'il est possible de tirer ces renseignements du type d'éclairage utilisé, cela évite des recherches supplémentaires (Nations Unies, 2008, par. 2.497).

644. Si la valeur est non valide, le code « inconnu » doit être attribué en l'absence d'imputation dynamique. Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, notamment, en règle générale, le type d'unité d'habitation, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

## 16. Évacuation des déchets solides : principal type (sujet de base)

645. D'après la publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* (Nations Unies, 2008, par. 2.500), ce sujet a trait à la collecte et l'évacuation des déchets solides produits par les occupants de l'unité d'habitation. L'unité de dénombrement est l'unité

d'habitation. Les directives pour la classification des unités d'habitation selon le mode d'évacuation des déchets solides sont présentées ci-dessous :

- 1 Déchets solides collectés périodiquement par des agents agréés
- 2 Déchets solides collectés de façon non périodique par des agents agréés
- 3 Déchets solides collectés par des agents autonomes
- 4 Déchets solides déposés par les occupants dans une décharge locale contrôlée par les autorités
- 5 Déchets solides déposés par les occupants dans une décharge locale non contrôlée par les autorités
- 6 Déchets solides incinérés par les occupants
- 7 Déchets solides enterrés par les occupants
- 8 Déchets solides déversés par les occupants dans un cours d'eau, en mer, dans un ruisseau ou dans un étang
- 9 Déchets solides transformés en compost par les occupants
- 10 Autres dispositions

646. Le type d'évacuation des déchets solides est indépendant des autres variables concernant l'habitation. Si la valeur est non valide, le code « inconnu » doit être attribué en l'absence d'imputation dynamique. Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, notamment, en règle générale, le type d'unité d'habitation, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation ou la présence d'installations pour la préparation des repas afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

## 17. Chauffage : type et énergie utilisée (sujet supplémentaire)

647. Ce sujet a trait au type de chauffage des unités d'habitation et à l'énergie consommée à cet effet. Les unités de dénombrement sont toutes les unités d'habitation. Ce sujet ne s'applique pas à un certain nombre de pays où, du fait de leur situation géographique et de leur climat, il n'est pas nécessaire de prévoir de chauffer les locaux d'habitation. Le mode de chauffage se réfère au type de système utilisé pour chauffer la majeure partie de l'espace habité. Ce peut être un chauffage central desservant toutes les unités d'habitation ou un ensemble de celles-ci, ou un chauffage autre que central, aménagé séparément dans un logement (poêle, cheminée ou autre installation). « L'énergie consommée pour le chauffage » est étroitement liée au mode de chauffage et désigne la principale source d'énergie, telle que combustibles solides (charbon, lignite et produits dérivés, bois), pétrole, combustibles gazeux (gaz naturel ou gaz liquéfié) et électricité (Nations Unies, 2008, par. 2.501).

648. Le mode de chauffage et l'énergie consommée pour le chauffage sont liés l'un à l'autre, ainsi qu'à l'alimentation en eau chaude et à d'autres services d'utilité publique utilisés dans l'unité d'habitation, tels que l'électricité et le gaz fourni par canalisations. Les équipes de vérification doivent tenir compte de la disponibilité de ces services lors de la mise au point des spécifications de vérification du mode de chauffage et de l'énergie consommée à cet effet. Le mode de chauffage peut être indépendant pour chaque unité d'habitation et peut avoir à être vérifié séparément. Toutefois, si la réponse à la question « énergie consommée pour le chauffage » est inconnue ou incohérente, le programme peut vérifier le type d'énergie utilisé pour l'éclairage. Enfin, si la valeur est non valide, le code « inconnu » doit être attribué en l'absence d'imputation dynamique. Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, notamment, en règle générale,

le type d'unité d'habitation, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

### **18. Eau chaude : disponibilité (sujet supplémentaire)**

649. Ce sujet a trait à la disponibilité de l'eau chaude dans les unités d'habitation. Par eau chaude, on entend une eau chauffée à une certaine température et acheminée par canalisations et robinets jusqu'aux occupants. Les informations recueillies peuvent indiquer si l'eau chaude est disponible dans les locaux d'habitation ou à l'extérieur de ceux-ci pour usage exclusif ou commun, ou si elle n'est pas disponible (Nations Unies, 2008, par. 2.502)

650. La disponibilité de l'eau chaude peut être liée au mode de chauffage de l'eau, quoique l'utilisation de l'énergie solaire à cet effet ne soit pas nécessairement liée à d'autres données sur l'habitation. Il incombe aux équipes de vérification de décider des vérifications appropriées, selon d'autres données sur l'habitation et la situation géographique. Enfin, si la réponse est non valide, le code « inconnu » doit être attribué en l'absence d'imputation dynamique. Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, telles que celles concernant l'alimentation en eau par canalisations, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

### **19. Gaz acheminé par conduites : disponibilité (sujet supplémentaire)**

651. Ce sujet a trait à la disponibilité du gaz accessible par conduite dans les unités d'habitation. Le gaz transporté par conduites se définit généralement comme du gaz naturel ou manufacturé distribué par conduites et dont la consommation est enregistrée. Ce sujet peut ne pas s'appliquer à un certain nombre de pays qui ne disposent pas de sources d'approvisionnement en gaz naturel ou n'ont pas de réseaux de conduites d'alimentation. (Nations Unies, 2008, par. 2.503).

652. Le gaz transporté par conduite est sans lien avec les autres données sur l'habitation, sauf pour certains types d'éclairage et comme combustible pour la cuisson. Il appartient aux équipes de vérification de décider du chemin de vérification approprié et d'effectuer les vérifications de cohérence. Si la réponse reste non valide ou incohérente, le code « inconnu » doit être attribué en l'absence d'imputation dynamique. Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, telles que celles concernant l'énergie utilisée pour le chauffage, le type de bâtiment, le type d'unité d'habitation, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

### **20. Utilisation d'une unité d'habitation (sujet supplémentaire)**

653. « L'utilisation d'une unité d'habitation » vise à répondre à la question de savoir si une unité d'habitation est utilisée entièrement aux fins d'habitation (logement) ou non. L'unité d'habitation peut servir de logement ou comme local commercial, atelier de fabrication ou à d'autres usages (Nations Unies, 2008, par. 2.504).

654. « L'utilisation d'une unité d'habitation » est indépendante de celle des autres unités d'habitation. Si la réponse reste non valide, le code « inconnu » doit être

attribué en l'absence d'imputation dynamique. Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, telles que le type d'unité d'habitation, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

## 21. Occupation par un ou plusieurs ménages (sujet de base)

655. La rubrique occupation par un ou plusieurs ménages est indépendante des autres rubriques concernant l'habitation. Si la valeur attribuée à cette variable est non valide, le pays doit alors compter les chefs de ménage et utiliser le nombre ainsi obtenu. Il convient de noter que cette vérification doit se faire après la vérification structurelle servant à identifier le chef de ménage.

## 22. Nombre d'occupants (sujet de base)

656. Chaque personne qui réside habituellement dans une unité d'habitation ou autres locaux collectifs à usage d'habitation doit être comptée comme occupant. L'unité de dénombrement pour ce sujet est donc le local d'habitation. Cependant, les recensements de l'habitation étant généralement effectués simultanément avec ceux de la population, cette définition sera plus ou moins applicable suivant que les renseignements recueillis et enregistrés pour chaque personne dans le recensement de la population indiquent où se trouvait cette personne le jour du recensement, ou se réfèrent à son lieu de résidence habituel. Il faut veiller à distinguer, parmi les occupants des unités d'habitation mobiles (bateau, roulotte, caravane, etc.) les personnes qui résident habituellement dans ces unités de celles qui les utilisent comme moyen de transport (Nations Unies, 2008, par. 2.510).

657. Le « nombre d'occupants » est lié au nombre d'enregistrements démographiques et ces deux nombres doivent être égaux. S'ils diffèrent, il faut alors corriger soit le nombre d'occupants, soit le nombre d'enregistrements concernant la population. D'ordinaire, le nombre d'occupants est ajusté de façon à être égal au nombre de personnes enregistrées dans l'unité d'habitation. Cette rubrique ne doit pas porter la mention « inconnu » et ne doit pas non plus faire l'objet d'une imputation.

## 23. Bâtiment : type (sujet de base)

658. La classification suivante par type de bâtiment est recommandée par les Nations Unies (2008, par. 2.514) pour les bâtiments dont une partie de l'espace est utilisée comme lieu d'habitation.

- 1 Bâtiments ne contenant qu'une unité d'habitation
  - 1.1 Maison individuelle
  - 1.2 Maison mitoyenne
- 2 Bâtiments contenant plus d'une unité d'habitation
  - 2.1 D'un maximum de 2 étages
  - 2.2 De 3 à 4 étages
  - 2.3 De 5 à 10 étages
  - 2.4 De 11 étages ou plus
- 3 Bâtiments pour personnes vivant en institution
- 4 Autres

659. Si la réponse est non valide ou incohérente, le code « inconnu » doit être attribué en l'absence d'imputation dynamique. Les pays qui optent pour l'imputation

dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, parmi lesquelles pourraient figurer les matériaux de construction des murs extérieurs, la période de construction et/ou le type d'unités d'habitation se trouvant dans le bâtiment, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

#### **24. Année ou période de construction (sujet supplémentaire)**

660. L'année ou la période de construction indique l'âge du bâtiment dans lequel est situé un ensemble de locaux d'habitation. Il est recommandé que soit obtenue l'année exacte de construction pour les bâtiments construits pendant la période intercensitaire précédant immédiatement le recensement si elle ne dépasse pas 10 ans. Si la période intercensitaire dépasse 10 ans ou s'il n'y a pas été effectué précédemment de recensement, il convient de chercher à obtenir l'année exacte de construction pour les bâtiments construits durant les 10 années précédentes. Pour les bâtiments construits depuis plus de 10 ans, il convient de recueillir des informations propres à faciliter la détermination de l'âge des bâtiments. Il peut s'avérer difficile d'obtenir de telles informations dans la mesure où les occupants peuvent ne pas connaître l'année de construction (Nations Unies, 2008, 2.519)

661. Certains pays, y compris parmi ceux qui utilisent l'imputation dynamique, acceptent la mention « inconnu » comme réponse concernant l'année ou la période de construction. Lorsque tel est le cas, le pays peut décider de ne pas utiliser l'imputation dynamique pour cette rubrique, même s'il emploie des matrices d'imputation pour d'autres variables. Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, parmi lesquelles pourraient figurer le type de bâtiment, les matériaux de construction des murs extérieurs et/ou le type d'unités d'habitation se trouvant dans le bâtiment, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

#### **25. Immeuble : Nombre de logements (sujet supplémentaire)**

662. La description des vérifications à effectuer concernant le nombre d'unités d'habitation d'un immeuble est donnée au chapitre III, dans le cadre des vérifications structurelles.

#### **26. Matériaux de construction des murs extérieurs (sujet de base)**

663. Ce sujet a trait aux matériaux de construction des murs extérieurs du bâtiment où se trouve l'ensemble d'unités d'habitation. Si ces murs sont construits à l'aide de plus d'un type de matériau, il convient d'enregistrer le matériau dominant. Les types de matériau signalés (briques, béton, bois, torchis) sont fonction des matériaux le plus souvent utilisés dans le pays en question et de leur importance du point de vue de la permanence de la construction et de l'évaluation de sa durabilité (Nations Unies, 2008, par. 2.525)

664. Si la réponse est non valide, en l'absence d'imputation dynamique, la mention « inconnu » doit être attribuée. Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, telles que la période de construction et/ou le type d'unités d'habitation se trouvant dans le bâtiment, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.



## 27. Matériaux de construction des sols, de la toiture (sujet supplémentaire)

665. Dans certains cas, les données sur les matériaux utilisés pour la construction de la toiture et des sols peuvent présenter un intérêt particulier et permettre une meilleure évaluation de la qualité des logements du bâtiment. Ce sujet a trait aux matériaux utilisés pour la toiture et/ou les sols (toutefois, selon les besoins particuliers du pays, il peut se référer également à d'autres parties du bâtiment, tels que la charpente ou les fondations). L'unité de dénombrement est ici le bâtiment. Seul le matériau dominant est recensé et, dans le cas de la toiture, ce peut être des tuiles, du béton ou des tôles métalliques, des branches de palmier, du chaume, du bambou ou autre matériau similaire, ou de la boue, des feuilles de plastique ou quelque autre matériau (Nations Unies, 2008, par. 2.528).

666. Parfois, la réponse donnée pour le matériau de construction des murs extérieurs ne concorde pas avec celle donnée pour le matériau de construction de la toiture; tel est le cas, par exemple, lorsque le matériau de construction indiqué par les murs n'est pas assez solide pour soutenir la toiture. Comme on l'a indiqué précédemment, il appartient aux spécialistes de décider de modifier l'une des deux variables ou d'utiliser la mention « inconnu ». Si une donnée est non valide, en l'absence d'imputation dynamique, la mention « inconnu » doit être attribuée. Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, telles que le type de bâtiment, le matériau de construction des murs extérieurs, le type d'unités d'habitation, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

667. Il se peut que le matériau de construction utilisé pour les sols soit ou non compatible avec ceux de la toiture et des murs. Si l'équipe nationale de vérification constate un défaut de concordance ou des combinaisons non valides, en l'absence d'imputation dynamique, elle doit décider d'attribuer la mention « inconnu ». Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, telles que le type de bâtiment, le matériau de construction des murs extérieurs, le type d'unités d'habitation, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

## 28. Ascenseur : disponibilité (sujet supplémentaire)

668. Ce sujet vise l'existence d'un ascenseur (plate-forme entourée de parois qui monte ou descend pour transporter des personnes et du fret) dans un immeuble à plusieurs étages. Les données se rapportent à un ascenseur utilisable la plupart du temps, c'est-à-dire opérationnel la plupart du temps, sous réserve d'entretien régulier (Nations Unies, 2008, par. 2.529).

669. Si l'immeuble n'a qu'un étage ou s'il s'agit d'une unité d'habitation individuelle, il ne doit pas y avoir d'ascenseur. S'il est fait état d'un ascenseur, l'équipe de vérification doit décider de l'information à considérer en priorité, soit le nombre d'étages, soit l'existence indiquée d'un ascenseur. Si l'existence d'un ascenseur est retenue en priorité, il faut alors modifier le nombre d'étages, soit en attribuant à cette rubrique le code « inconnu », soit en obtenant une autre valeur par imputation dynamique. Si le nombre d'étages est considéré en priorité, et si l'immeuble ne comporte qu'un seul niveau, la réponse à la question « existence d'un ascenseur » doit être « non ».

670. Si l'existence d'un ascenseur est indiquée et s'il fonctionne à l'électricité, il convient de faire une vérification pour s'assurer que l'immeuble est bien relié au secteur.

671. Enfin, si la valeur indiquée pour la rubrique « existence d'un ascenseur » n'est pas valide, en l'absence d'imputation dynamique, il convient de lui attribuer la mention « inconnu ». Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, telles que le type de bâtiment et le matériau de construction des murs extérieurs, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

## **29. Bâtiment à usage agricole (sujet supplémentaire)**

672. Certains pays jugent utile pour leurs recensements nationaux de préciser si un immeuble recensé est ou non à usage agricole. Par immeuble à usage agricole, on entend un immeuble qui fait partie d'une exploitation agricole et est utilisé à des fins agricoles et/ou d'habitation (Nations Unies, 2008, par. 2.531).

673. La rubrique « immeuble à usage agricole » est indépendante des autres rubriques concernant l'habitation. Les pays peuvent choisir de vérifier la concordance avec les données sur la population concernant la profession et le secteur d'activité. Si la valeur indiquée est non valide, en l'absence d'imputation dynamique, il convient de lui attribuer la mention « inconnu ». Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

## **30. État d'entretien (sujet supplémentaire)**

674. Ce sujet sert à préciser si l'immeuble a besoin de réparations et à indiquer la nature des réparations nécessaires. L'unité de dénombrement est l'immeuble. La classification des immeubles selon leur état d'entretien peut inclure les catégories suivantes : « aucune réparation nécessaire », « petites réparations nécessaires », « réparations moyennes nécessaires » ou encore « grosses réparations nécessaires » et « irréparable ». Par « petites réparations », on entend principalement l'entretien ordinaire de l'immeuble et de ses composants (par exemple, vitres cassées). Les réparations moyennes correspondent à la rectification de défauts d'importance moyenne comme l'absence de gouttières sur le toit, des plâtres endommagés sur de grandes surfaces ou des rampes peu sûres aux escaliers. Les grosses réparations sont nécessitées par un vice de construction de l'immeuble, tel que tuiles ou bardeaux manquants au toit, fissures et trous dans les murs extérieurs et marches manquantes. Le terme « irréparable » s'applique aux immeubles qui ne peuvent plus être réparés, les défauts structurels étant tellement graves qu'il est jugé préférable de démolir l'immeuble plutôt que d'entreprendre de le réparer. Ce terme désigne le plus souvent les immeubles dont il ne reste que la charpente et qui n'ont plus de murs extérieurs ni de toit (Nations Unies, 2008, par. 2.532).

675. L'état d'entretien de l'immeuble est indépendant des autres variables concernant l'habitation. Par conséquent, si la valeur indiquée est non valide, en l'absence d'imputation dynamique, il convient de lui attribuer la mention « inconnu ». Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

### **31. Caractéristiques du chef de ménage ou d'un autre membre du ménage utilisé comme référence (sujet de base).**

676. Les caractéristiques du chef de ménage sont généralement tirées des enregistrements concernant la population, afin de faciliter l'établissement des informations qui seront mises en tableaux croisés à des fins de planification et d'analyse. Ces rubriques, notamment le sexe, l'âge, l'origine ethnique, la religion ou le revenu, facilitent la détermination des différences de statut social ou des besoins sociaux. Les caractéristiques en question ayant déjà fait l'objet d'une vérification concernant les rubriques démographiques, elles n'ont pas à être soumises à une vérification supplémentaire dans le présent contexte (Nations Unies, 2008, par. 2.533).

### **32. Modalités de jouissance (sujet de base)**

677. D'après les Nations Unies (2008, par. 2.536), les modalités de jouissance indiquent à quel titre un ménage occupe tout ou partie d'une unité d'habitation. L'unité de dénombrement est le ménage occupant une unité d'habitation. La classification des ménages selon les modalités de jouissance est la suivante :

- 1 Membre du ménage propriétaire de l'unité d'habitation
- 2 Membre du ménage louant tout ou partie de l'unité d'habitation
  - 2.1 Membre du ménage louant tout ou partie de l'unité d'habitation en tant que locataire en titre
  - 2.2 Membre du ménage louant tout ou partie de l'unité d'habitation en tant que sous-locataire
- 3 Unité d'habitation occupée sans paiement de loyer
- 4 Autre arrangement.

678. La rubrique modalités de jouissance peut avoir un lien avec le type de propriété; aussi, l'équipe de vérification doit-elle parfois envisager le lien entre ces deux rubriques. Sinon, en cas de non-validité de la valeur indiquée pour la variable modalités de jouissance, en l'absence d'imputation dynamique, il convient de lui attribuer le code « inconnu ». Les pays qui optent pour l'imputation dynamique de valeurs non valides doivent s'appuyer au moins sur deux caractéristiques pour obtenir des informations « connues » auprès d'unités d'habitation analogues de la même zone géographique.

### **33. Loyer et coût de logement des propriétaires occupants (sujet supplémentaire)**

679. La rubrique relative aux loyers et aux coûts de logement des propriétaires occupants est indépendante des autres variables concernant l'habitation, si ce n'est évidemment que les loyers se rapportent uniquement aux unités occupées par des locataires et les coûts de logement exclusivement aux unités occupées par leurs propriétaires. L'équipe de vérification doit examiner chaque cas et déterminer les liens les plus appropriés entre ces variables. Si la valeur indiquée est non valide, en l'absence d'imputation dynamique, il convient de lui attribuer la mention « inconnu ». Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique (Nations Unies, 2008, par. 2.540).

### **34. Unités d'habitation meublées ou non meublées (sujet supplémentaire)**

680. Cette rubrique, qui traite de la question de savoir si les unités d'habitation sont meublées ou non meublées est nouvelle. Les équipes de vérification doivent envisager de la soumettre à vérification, si elle est incluse, afin de déterminer les meilleures rubriques à utiliser pour l'imputation dynamique, si cette méthode est utilisée pour résoudre les cas de réponses non valides ou incohérentes (Nations Unies, 2008, par. 2.542).

### **35. Dispositifs d'application des technologies de l'information et de la communication : disponibilité (sujet de base)**

681. Les dispositifs d'application des technologies de l'information et de la communication (TIC) prennent de plus en plus d'importance dans la société contemporaine. Ces dispositifs offrent un ensemble de services qui sont en train de transformer la structure et la configuration des principaux phénomènes sociaux et économiques. Le recensement de l'habitation offre une possibilité unique d'évaluer la disponibilité de ces dispositifs au sein des ménages. Le choix des sujets devrait être suffisant pour permettre de comprendre la place des TIC dans le ménage, ainsi que son utilisation aux fins de planification par les secteurs public et privé, de manière à faciliter une prestation plus large et meilleure des services, et à évaluer leur impact sur la société. La classification recommandée est la suivante :

- 1 Ménage ayant la radio
- 2 Ménage ayant un poste de télévision
- 3 Ménage ayant une ligne téléphonique fixe
- 4 Ménage ayant un/des téléphone/s portable/s cellulaire/s
- 5 Ménage ayant un/des ordinateur/s individuel/s
- 6 Ménage ayant accès à Internet depuis son domicile
- 7 Ménage ayant accès à Internet depuis un autre lieu que son domicile
- 8 Ménage n'ayant pas accès à Internet

682. Les dispositifs d'application des technologies de l'information et de la communication sont de nouveaux éléments. Les dispositifs alimentés à l'électricité ne devraient être disponibles que dans les immeubles reliés au secteur. Toutefois, à mesure que l'utilisation de l'énergie solaire, de l'énergie éolienne et d'autres sources d'énergie « renouvelables » devient plus fréquente, il faut en tenir compte lors de la conception de programmes de vérification pour ces équipements. Les équipes nationales de vérification doivent procéder à une vérification minutieuse de ce sujet et de ses matrices d'imputation avant d'entreprendre le recensement ou l'enquête. Parmi les rubriques qu'il serait utile de prévoir pour l'imputation dynamique figurent le niveau social du ménage (déterminé, par exemple, à l'aide de l'indice de richesse) et de l'âge du chef de ménage (Nations Unies, 2008, par. 2.543).

683. Ces questions concernent la disponibilité d'un dispositif dans l'unité d'habitation. Par exemple, la mention d'un téléphone implique l'existence d'une ligne téléphonique plutôt que la présence concrète d'un téléphone particulier, car plusieurs téléphones peuvent être branchés sur une même ligne téléphonique (Nations Unies, 2008, par. 2.547 et 2.548). Les téléphones ne sont pas liés à d'autres rubriques concernant l'unité d'habitation lors de la vérification. Toutefois, si certaines zones géographiques n'ont pas le téléphone, l'équipe de vérification doit en tenir compte lors de la mise au point des programmes de vérification. Si la valeur indiquée pour le « téléphone »

n'est pas valide, en l'absence d'imputation dynamique, il convient d'attribuer la mention « inconnu ». Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, telles que le type d'unité d'habitation, les matériaux de construction des murs et les modalités de jouissance, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

### **36. Automobiles : nombre (sujet supplémentaire)**

684. Le « nombre de voitures » fait référence au nombre de voitures et camionnettes dont disposent normalement les membres du ménage, c'est-à-dire dont ils sont propriétaires ou qui font l'objet d'un accord plus ou moins permanent comme un contrat de location, ou qui leur sont fournis par un employeur s'ils en ont l'usage personnel, mais ne comprend pas les camionnettes utilisées exclusivement pour le transport de marchandises (Nations Unies, 2008, par. 2.551).

685. La variable « nombre de véhicules » est indépendante des autres variables concernant l'habitation. Si dans certaines zones du pays, il n'y a pas de voitures, les spécialistes peuvent envisager de mettre au point des vérifications spéciales adaptées à des zones géographiques particulières. Sinon, si la valeur indiquée n'est pas valide, en l'absence d'imputation dynamique, il convient de lui attribuer la mention « inconnu ». Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, telles que le type d'unité d'habitation, les matériaux de construction des murs, les modalités de jouissance et le type de propriété ou encore, dans ce cas particulier, le nombre d'occupants adultes, afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

### **37. Équipements ménagers : disponibilité (sujet supplémentaire)**

686. Les informations recueillies concernent la présence éventuelle d'équipements ménagers tels que machine à laver, lave-vaisselle, réfrigérateur, congélateur, etc. selon le contexte national (Nations Unies, 2008, par. 2.552).

687. Le fonctionnement de la plupart des appareils en question exigent que l'unité d'habitation considérée ait l'électricité. Lorsque le questionnaire fait état de ces équipements, le programme de vérification mis au point doit contrôler la disponibilité de l'électricité dans ces locaux (à l'exception, le cas échéant, d'un réfrigérateur fonctionnant au gaz ou d'une glacière). Si dans le pays considéré, le lave-linge ou le lave-vaisselle doit fonctionner à l'eau courante, cette exigence doit être prise en compte dans le programme de vérification. Il est également possible, par le biais de certains contrôles, de vérifier si la présence d'un élément particulier doit ou non être fonction de la disponibilité de l'électricité et de l'eau et d'effectuer les corrections nécessaires si des incohérences sont constatées. Là encore, certaines zones géographiques d'un pays sont parfois dépourvues d'électricité ou d'eau courante; dans ce cas, les spécialistes doivent en tenir compte lors de la mise au point de leurs vérifications. Si la valeur indiquée est non valide ou incohérente, en l'absence d'imputation dynamique, il convient de lui attribuer la mention « inconnu ». Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, telles que le type d'unité d'habitation, l'électricité, les matériaux de construction des murs et les modalités de jouissance afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique (où les niveaux sociaux des ménages devraient être comparables).

### 38. Espaces extérieurs : disponibilité (sujet supplémentaire)

688. Ce sujet se réfère à l'existence éventuelle d'un espace extérieur prévu pour les loisirs des membres d'un ménage occupant une unité d'habitation. La classification vise l'espace extérieur faisant partie de l'unité de logement (par exemple, le jardin dans le cas d'une maison individuelle), l'espace extérieur proche du bâtiment (par exemple, cour ou terrain de jeux situé près de l'immeuble), un espace commun de loisirs situé à moins de 10 minutes de marche de l'unité d'habitation (par exemple, jardin public, centre sportif, etc.) ou des espaces situés à plus de 10 minutes à pied (Nations Unies, 2008, par 2.553).

689. L'espace extérieur dont disposent les membres du ménage est indépendant des autres rubriques concernant l'habitation. Toutefois, dans certaines zones géographiques ou pour certains types d'immeubles, l'espace extérieur est parfois inexistant. Les équipes de vérification peuvent avoir à envisager le contexte particulier lorsqu'elles établissent leurs programmes de vérification. Si la valeur indiquée est non valide, en l'absence d'imputation dynamique, il convient de lui attribuer la mention « inconnu ». Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer au moins sur deux caractéristiques, telles que le type de bâtiment ou le type d'unité d'habitation afin de recueillir des données « connues » à partir d'unités d'habitation similaires de la même zone géographique.

## B. Unités d'habitation occupées et inoccupées

690. Les vérifications décrites ci-dessus se rapportent aux unités d'habitation occupées. Toutefois, les unités d'habitation inoccupées et occupées présentent souvent des caractéristiques différentes et ne feront pas l'objet des mêmes vérifications. Aussi l'équipe de vérification du bureau national de recensement/statistique devra-t-elle mettre au point des vérifications différentes pour chaque type d'unité si, comme c'est généralement le cas, toutes les données concernant l'habitat ne sont pas recueillies pour les unités d'habitation inoccupées. L'équipe de vérification devra procéder avec le plus grand soin au choix des variables de la matrice d'imputation, celles-ci étant très vraisemblablement différentes.

691. Le présent chapitre a été consacré à l'examen des variables concernant l'habitat recommandées dans la publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision*. Aucun pays ne devrait utiliser toutes ces variables, et les variables choisies ainsi que leurs relations géographiques avec les autres variables devraient être soigneusement testées en vase clos et dans les situations observées avant le recensement afin de s'assurer que les réponses obtenues sont complètes et fiables. Les variables concernant l'habitat sont assez importantes par elles-mêmes pour être utilisées dans l'établissement d'une partie de l'indice de richesse afin d'évaluer le bien-être de la population dans toutes les régions du pays.

## Annexe I

# Variables calculées

1. Afin de tirer le meilleur parti des données recueillies à l'occasion des recensements ou des enquêtes, les pays ont fréquemment besoin d'utiliser des variables qui sont elles-mêmes des combinaisons d'autres variables. Par exemple, les informations concernant la situation de l'activité économique (voir chapitre IV, section D.1) constituent d'ores et déjà une combinaison de plusieurs variables recueillies lors du recensement. Au lieu de devoir élaborer un programme de recodage, chaque fois que le bureau national de recensement/statistique veut obtenir un tableau particulier, les informaticiens peuvent rédiger un programme permettant d'effectuer le recodage une fois pour toutes, de stocker les informations recodées dans l'enregistrement personnel considéré, puis de les utiliser pour établir d'autres tableaux. Les bureaux nationaux de recensement/statistique doivent évaluer la fréquence d'utilisation d'informations recodées, ainsi que la pertinence d'une information recodée particulière lorsqu'ils décident de produire et d'enregistrer l'information. Par ailleurs, il ne faut pas oublier que les informations recodées occupent de la place dans les enregistrements personnels. L'espace nécessaire à cet effet sera d'autant plus considérable que la population sera nombreuse.

2. Il est possible de créer de cette façon un grand nombre de variables. Par exemple, si la date de naissance est indiquée mais non l'âge, celui-ci peut être obtenu en une seule opération en soustrayant la date de naissance de la date de référence du recensement, et cette information sera stockée dans l'enregistrement. De même, on peut obtenir le revenu du ménage en additionnant les revenus de chaque individu et en plaçant la somme calculée dans l'enregistrement relatif à l'habitation en vue d'une utilisation ultérieure.

3. Parfois, les variables calculées peuvent provenir de la combinaison d'une ou plusieurs données figurant dans un enregistrement particulier ou parfois dans plusieurs enregistrements. Par exemple, la classification dans la catégorie « population inactive : fréquente l'école » peut exiger l'examen des réponses fournies à quatre questions. Ainsi, lors de la définition de formats de tableaux ou de la conception de tableaux supplémentaires, l'utilisation de variables calculées facilitera la programmation et augmentera son efficacité et contribuera à améliorer la comparabilité des données dans le temps. On trouvera ci-dessous quelques exemples d'enregistrements calculés.

### A. Variables calculées associées aux données sur l'habitation

#### 1. Revenu du ménage

4. La variable calculée pour le revenu du ménage est égale à la somme des revenus de toutes les catégories de revenus de chacun des membres d'un ménage. Parmi les différentes catégories de revenus figurent les salaires, les revenus d'activités d'entreprises, les intérêts et dividendes, les prestations de sécurité sociale, les retraites, les envois

de fonds, les rentes et les loyers. Si l'on veut obtenir le revenu total, il convient, dans le cadre de la vérification, de vérifier le revenu total de chaque personne en additionnant les catégories individuelles. On vérifie ensuite ce total par rapport au revenu total indiqué. Si le revenu cumulé obtenu est différent du revenu total déclaré, les équipes de vérification doivent mettre au point une correction. Il faut soit modifier le total, de façon à le faire correspondre à la somme des parties, soit modifier les chiffres concernant une ou plusieurs catégories individuelles. Une fois fixés les revenus totaux pour chacun de ses membres, on obtient la variable correspondant au revenu du ménage en additionnant les revenus individuels.

5. L'équipe de vérification doit tenir compte de l'éventualité d'un revenu devenu négatif pour un ou plusieurs membres du ménage à la suite d'une faillite économique ou d'une autre cause. Dans ce cas, le revenu total du ménage sera réduit et non augmenté du « revenu » de ce membre.

## 2. Revenu familial

6. La variable calculée pour le revenu familial est égale à la somme des revenus de toutes les catégories de revenus de chacun des membres de la famille. Contrairement aux ménages, les familles sont d'ordinaire uniquement constituées d'individus apparentés, bien que cette définition soit propre à la situation particulière au pays considéré. Dans certains pays, ménages et famille sont identiques, de telle sorte qu'une variable calculée correspondant au revenu familial s'avère inutile. Parmi les différentes catégories de revenus peuvent figurer les salaires, les revenus d'activités commerciales, les intérêts et dividendes, les revenus des prestations sociales, les retraites, les envois de fonds, les rentes et les loyers. Si l'on veut obtenir le revenu total, il convient, lors de la vérification, de vérifier le revenu total de chaque personne en additionnant les catégories individuelles. On vérifie ensuite ce total par rapport au revenu total déclaré. Si le revenu cumulé obtenu est différent du revenu total déclaré, les équipes de vérification doivent mettre au point une correction. Il faut soit modifier le total, de façon à le faire correspondre à la somme des parties, soit modifier les chiffres concernant une ou plusieurs catégories individuelles. Une fois fixés les revenus totaux pour chacun de ses membres, on obtient la variable correspondant au revenu familial en additionnant les revenus individuels.

7. L'équipe de vérification doit tenir compte de l'éventualité d'un revenu devenu négatif pour un ou plusieurs membres de la famille à la suite d'une faillite économique ou d'une autre cause. Dans ce cas, le revenu total de la famille sera réduit et non augmenté du « revenu » de ce membre.

## 3. Noyau familial

8. Pour la composition des ménages, la publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitation, deuxième révision* a établi un code pour le noyau familial, défini comme l'une des données recodées ci-dessous, avec suggestion de recodage présentée entre parenthèses :

1. Couple marié (ou couple vivant en union consensuelle) sans enfants (chef de ménage et conjoint, cochefs de ménage ou couple vivant en union consensuelle)
2. Couple marié (ou couple vivant en union consensuelle) avec un ou plusieurs enfants non mariés (comme dans le cas précédent, mais par la recherche du ménage ou l'établissement d'une information recodée pour le nombre d'enfants non mariés vivant dans l'unité d'habitation, avec au moins un enfant non marié)



3. Père avec un ou plusieurs enfants non mariés (chef de ménage du sexe masculin, pas d'épouse présente, et au moins un enfant non marié, déterminé comme il est indiqué ci-dessus)
4. Mère avec un ou plusieurs enfants non mariés (chef de ménage du sexe féminin, pas de mari présent, et au moins un enfant non marié, déterminé comme il est indiqué ci-dessus)

#### 4.. Type de famille

9. La publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitat, deuxième révision* présente une description générale de divers types de familles afin d'aider à établir un code pour la composition d'une famille. Les pays peuvent choisir de n'utiliser qu'un seul code, ou d'utiliser une série de codes selon l'utilisation éventuelle des données.

10. Un premier code pourrait servir à identifier le type de famille, représenté par les éléments suivants, avec définitions. Les codes suggérés sont indiqués dans la section suivante.

1. Ménage constitué d'une seule personne
2. Famille nucléaire : noyau familial, comprenant un couple marié ou des partenaires vivant en union consensuelle avec ou sans enfants ou parent vivant seul avec enfant(s)
3. Famille élargie : noyau familial et autres personnes apparentées au chef de ménage; deux noyaux familiaux ou plus; ou deux personnes ou plus apparentées l'une à l'autre mais ne faisant pas partie du noyau familial
4. Ménage composite (autres types de ménages)

#### 5. Composition des ménages

11. Les ménages constitués d'une seule personne sont des ménages plutôt que des familles et doivent donc être comptés comme catégorie distincte dans le recodage de la composition des ménages.

12. *Ménages composés d'une famille nucléaire.* Les ménages composés d'une famille nucléaire peuvent se subdiviser (avec codes individuels) en : 1) familles composées d'un couple marié avec enfants; 2) familles composées d'un couple marié sans enfants; 3) partenaires vivant en union consensuelle avec enfants; 4) partenaires vivant en union consensuelle sans enfants; 5) pères avec enfants; et 6) mères avec enfants. Pour déterminer le code approprié, on établit le sexe du chef de ménage, puis l'on recherche la présence d'un conjoint et d'enfants au sein du ménage. Le type de code pour ménages composés d'une famille nucléaire peut être un code à deux chiffres avec le chiffre 2 comme premier chiffre (le chiffre 1 étant réservé pour ménages constitués d'une seule personne); ainsi, le code 21 représenterait une famille composée d'un couple marié avec enfants.

13. *Ménages composés d'une famille élargie.* Les familles élargies peuvent également se diviser en catégories qui comprendraient (sur la base des désignations précédentes) : (31) un noyau familial unique et d'autres personnes apparentées aux membres de ce noyau; (32) deux noyaux familiaux ou plus apparentés l'un à l'autre sans autres personnes; (33) deux noyaux familiaux ou plus apparentés l'un à l'autre plus d'autres personnes apparentées aux membres de ces noyaux; et (34) deux personnes ou plus apparentées l'une à l'autre, dont aucune n'appartient à un noyau familial. On déterminerait les codes eux-mêmes en recherchant au sein du ménage les nombres de noyaux et les liens entre leurs membres et le ménage. Si le ménage est déjà codé comme nucléaire, cette procédure ne sera pas appliquée.

14. *Ménages composites.* Tous les autres ménages seraient des ménages composites. L'utilisation de la même procédure que précédemment donnerait les résultats suivants : (41) un noyau familial unique plus d'autres personnes, dont certaines sont apparentées aux membres de ce noyau et d'autres ne le sont pas; (42) un noyau familial unique plus d'autres personnes, dont aucune n'est apparentée aux membres de ce noyau; (43) deux noyaux familiaux ou plus apparentés l'un à l'autre, plus d'autres personnes, dont certaines sont apparentées au moins aux membres de l'un de ces noyaux et d'autres ne sont apparentés à aucun de ces noyaux; (44) deux noyaux familiaux ou plus apparentés l'un à l'autre, plus d'autres personnes, dont aucune n'est apparentée à aucun de ces noyaux; (45) deux noyaux familiaux ou plus non apparentés l'un à l'autre, avec ou sans autres personnes; (46) deux personnes ou plus apparentées les unes aux autres, dont aucune n'appartient à un noyau familial, plus d'autres personnes non apparentées; et (47) personnes non apparentées. Là encore, une série de recherches et des résumés permettront de choisir la désignation appropriée à attribuer à chaque type de ménage.

## 6. Composition de la famille

15. Les familles sont une subdivision des ménages, de sorte que le nouveau code de composition de la famille comprendra les catégories appropriées pour les familles décrites ci-dessus. Dans la mesure où un ménage à une personne ne constitue pas une famille, il ne figurera pas dans le nouveau code de composition de la famille. De même, les ménages composites sont des ménages mais pas des familles, de sorte qu'ils ne seront pas non plus inclus. Chaque pays décidera donc ou non d'inclure un nouveau code unique pour toutes les familles (nucléaires et élargies) ou des codes distincts pour ces deux types de familles, étant entendu que ces nouveaux codes ne se recouperont pas (on pourrait toutefois envisager de grouper les ménages composés d'une famille nucléaire avec les familles élargies).

## 7. Ménage et état matrimonial

16. Le ménage et l'état matrimonial indiquent le lien qui existe entre une personne et des autres membres du ménage ou de la famille. La méthode suivie pour définir le ménage et l'état matrimonial diffère de la méthode traditionnelle de classification des membres d'un ménage uniquement en fonction de leur lien avec le chef de ménage ou la personne de référence.

17. La publication *Principes et Recommandations concernant les recensements de la population et de l'habitat, deuxième révision* suggère le programme de codage suivant pour définir l'état matrimonial. Le premier ensemble de codes a trait aux personnes composant les ménages constitués d'au moins un noyau familial (c'est-à-dire les ménages qui sont aussi des familles). Les nouveaux codes suggérés sont notamment :

- 1.1 Mari (chef de ménage ou conjoint du sexe masculin)
- 1.2 Femme (chef de ménage ou conjoint du sexe féminin)
- 1.3 Partenaire d'une union consensuelle ou partenaire partageant l'habitation (à partir des codes de liens, le cas échéant, ou d'une combinaison de codes de liens et d'état matrimonial)
- 1.4 Mère célibataire (déterminé sur la base de l'absence de mari, mais de la présence d'enfants)
- 1.5 Père célibataire (déterminé sur la base de l'absence de femme, mais de la présence d'enfants)
- 1.6 Enfants vivant avec leurs deux parents (enfant du chef de ménage, avec présence des deux parents dans l'habitation)

- 1.7 Enfant vivant avec sa mère célibataire (enfant du chef de ménage, mais absence du père de l'enfant)
- 1.8 Enfant vivant avec son père célibataire (enfant du chef de ménage, mais absence de la mère de l'enfant)
- 1.9 Personne n'étant pas membre d'un noyau familial (tout autre parent). La publication *Principes et Recommandations concernant les recensements de la population et de l'habitat, deuxième révision* divise cette catégorie en deux groupes : 1) vivant avec des parents; et 2) vivant avec des personnes non apparentées

18. Le second ensemble de codes s'applique aux personnes d'un ménage sans noyau familial : personnes vivant seules; et personnes vivant avec d'autres parents et/ou avec des personnes non apparentées, à l'exclusion du conjoint ou d'un enfant du chef de ménage. Ces catégories sont les suivantes :

- 2.1 Personnes vivant seules (ménage composé d'une seule personne)
- 2.2 Personnes vivant avec d'autres personnes (personne vivant dans une unité d'habitation sans le conjoint ou un enfant du chef de ménage). Cette catégorie se subdivise elle-même en : 1) avec frères et sœurs; 2) avec des parents autres que frères et sœurs; ou 3) avec des personnes non apparentées

19. Une variable unique devrait être établie à partir de ces catégories, car celles-ci s'excluent mutuellement. Cette variable serait un code à deux chiffres. Certains bureaux de statistique peuvent désirer que le premier chiffre soit indépendant du second; dans ce cas, le premier chiffre indiquera si le ménage est ou non une famille nucléaire, et le second identifiera le type d'état matrimonial d'une personne donnée.

20. La publication *Principes et Recommandations concernant les recensements de la population et de l'habitat, deuxième révision* prévoit également des catégories pour la classification des personnes selon leur état matrimonial. Ces catégories comprennent : 1) homme ou femme d'un couple formant un ménage avec ou sans enfants; 2) parent célibataire, par sexe; 3) enfant du chef de ménage, enfant d'un couple marié ou enfant d'un parent célibataire, selon le sexe du parent; et 4) non-membre du noyau familial (apparenté ou non, et si apparenté, comment). Les définitions données ci-dessus pour la situation au sein d'un ménage sont applicables également à la situation au sein d'une famille.

## 8. Impact du VIH/sida sur la structure du ménage

21. Étant donné l'impact de l'épidémie de VIH/sida sur la structure des ménages de tant de pays, un nouveau code peut aider à décrire les différents types d'unités d'habitation. Par exemple, un code décrivant les ménages où une génération a disparu (et où ne restent que grands-parents et petits-enfants); les ménages dont le chef a moins de 18 ans; les ménages ayant une veuve pour chef de ménage, et ainsi de suite, peut être utilisé pour évaluer l'impact social et économique de l'épidémie, encore qu'indirectement. Les enfants qui entrent et sortent de la population active, la structure de la main-d'œuvre à l'intérieur des ménages, etc., peuvent aider les planificateurs à décrire pleinement l'impact du VIH/sida.

## 9. Personnes apparentées

22. Les personnes apparentées sont les personnes qui ont d'une certaine manière un lien avec le chef de ménage. La variable calculée pour les personnes apparentées est la somme de toutes les personnes liées au chef de ménage. Cette valeur est

particulièrement importante dans les situations où de nombreuses personnes non apparentées vivent ensemble dans des unités d'habitation. Lorsque de nombreuses personnes non apparentées vivent ainsi, elles sont souvent classées comme vivant dans des « locaux d'habitation collectifs » ou dans des « logements collectifs ».

23. Lors de la création de jeux de données, les bureaux nationaux de statistique définissent souvent des variables calculées par des ensembles différents de personnes apparentées, regroupées par tranche d'âge. Par exemple, des variables calculées peuvent être définies pour les enfants apparentés âgés de 0 à 5 ans, de 5 à 17 ans, de 6 à 17 ans ou de 0 à 17 ans, ou encore pour des personnes de 65 ans et plus ou de 75 ans et plus.

24. Les « enfants apparentés » pourraient comprendre par exemple les propres enfants du chef de ménage, ainsi que les personnes de moins de 18 ans, indépendamment de leur état matrimonial, qui sont apparentées au chef de ménage, à l'exception de son conjoint. Parmi les enfants apparentés pourraient ou non figurer les enfants adoptifs, ces derniers n'étant pas apparentés au chef de ménage, cette décision étant toutefois fonction du contexte national particulier.

## **10. Travailleurs de la famille**

25. Les pays souhaitent parfois comparer les variables concernant le ménage en fonction du nombre de travailleurs, par exemple, les distributions du revenu en fonction de la taille du ménage, et les nombres de travailleurs par personne à charge. Le pays peut ainsi déterminer la variable calculée correspondant au nombre de travailleurs de la famille, en additionnant le nombre de personnes ayant travaillé au moins une heure au cours d'une période de référence, par exemple une semaine ou un an (année civile ou les douze derniers mois). Ainsi, le nombre de personnes ayant travaillé au cours de la semaine passée pourrait être utilisé si les données recueillies concernent exclusivement cette période.

## **11. Installations complètes de plomberie**

26. Plusieurs rubriques du questionnaire de recensement permettent de recueillir des données sur les installations de plomberie. Ces rubriques sont généralement associées à la présence de canalisations d'eau, d'un cabinet d'aisances à chasse d'eau, d'une baignoire ou d'une douche, et ces données sont généralement recueillies auprès des unités d'habitation tant occupées qu'inoccupées. Une variable calculée associée aux installations complètes de plomberie peut faciliter la comparaison de la situation socioéconomique de différents groupes à un moment donné, ou son évolution dans le temps. Ce type de variable calculée pourrait être obtenu par exemple lorsque 3 types d'équipement, conduites d'eau (chaude et froide), lieu d'aisances à chasse d'eau et baignoire ou douche, sont présents (soit à l'intérieur de l'unité d'habitation, soit à l'extérieur de l'immeuble dans lequel se trouve l'unité d'habitation). L'équipe de vérification devra déterminer le jeu de variables le plus approprié concernant l'installation complète de plomberie.

27. Dans cet exemple, la variable calculée peut être obtenue lorsque la collecte de données porte séparément sur les trois rubriques et, pendant la vérification, la présence simultanée de ces trois équipements est établie. Si l'unité d'habitation est munie de canalisations d'eau, d'un lieu d'aisances à chasse d'eau et d'une baignoire ou d'une douche, elle est alors équipée d'une installation complète de plomberie. En l'absence de ces trois éléments, elle est dépourvue de telles installations.

## 12. Installation de cuisines complètes

28. Les recensements sont mis à profit pour recueillir des données sur les installations de cuisines complètes à partir des rubriques du questionnaire concernant ce type d'équipements, ainsi que les réfrigérateurs et les éviers; ces données sont recueillies pour les unités d'habitation tant occupées qu'inoccupées. Ainsi, une unité d'habitation peut être considérée comme étant équipée d'installations de cuisines complètes quand des installations de préparation des aliments (fourneau électrique, au kérosène ou au gaz, four micro-ondes et foyer non portatif ou cuisinière), un réfrigérateur et un évier relié à des canalisations d'eau se trouvent dans le même immeuble que les locaux d'habitation recensés. Ils ne doivent pas nécessairement se trouver dans la même pièce.

29. La variable calculée s'obtient quand les trois éléments d'information sont recueillis séparément et quand la vérification permet de conclure à la présence de ces trois éléments. « L'absence d'installation de cuisines complètes » correspond à une situation dans laquelle les trois équipements spécifiés sont présents, mais ne se trouvent pas dans le même immeuble; certains seulement de ces équipements sont présents; ou encore, aucun des trois équipements spécifiés ne se trouve dans le même immeuble que les locaux d'habitation recensés.

## 13. Loyer brut

30. Les pays peuvent recueillir des données sur les loyers réglés en espèces ou les loyers charges comprises. En général, les loyers payés en espèces ne comprennent pas les charges. Parfois, les pays ont besoin également de recueillir des informations sur le loyer brut. Le loyer brut est défini comme le loyer net augmenté du coût mensuel estimé des charges (électricité, gaz et eau) et des combustibles utilisés (notamment le mazout, le charbon, le kérosène et le bois) si les paiements correspondants sont à la charge du locataire. Le loyer brut est censé éliminer les différences résultant de diverses pratiques quant à l'inclusion des charges et combustibles dans le règlement du loyer; les unités d'habitation louées occupées sans paiement de loyer peuvent figurer séparément dans les tableaux sous une rubrique « sans loyer ».

31. La variable calculée relative au loyer brut s'obtient en additionnant le montant du loyer et celui des charges, si ces dernières sont payées séparément.

## 14. Indice de richesse

32. L'indice de richesse mesure le bien-être dans un pays ou dans certaines régions d'un pays. Le plus souvent, cet indice s'établit à partir des biens du ménage. Souvent, on utilise une analyse factorielle pour obtenir le meilleur ensemble de rubriques et leurs variantes. Généralement, on attribue des valeurs binaires à ces rubriques — 1 pour « présent » et 0 pour « absent » — puis on additionne les résultats. Plus le chiffre est élevé, plus il y a de richesse. Par exemple, le fait de posséder un téléviseur serait codé 1 (présence) ou 0 (absence). Par ailleurs, un cabinet d'aisances pourrait être codé 1 pour « situé à l'extérieur », 2 pour « chasse d'eau par gravité » ou 3 pour « chasse d'eau » (ce qui implique trois ensembles de variables binaires). Les divers éléments pourraient être ensuite pondérés lors de l'établissement de leur somme.

33. On peut ensuite établir des quintiles en prenant chaque cinquième de la distribution des valeurs de l'indice de richesse. Le quintile le plus bas correspondrait aux ménages les plus pauvres, et le quintile le plus haut aux ménages les plus riches.

## B. Variables calculées associées aux données sur la population

### 1. Situation de l'activité économique

34. Le fait de disposer d'une variable calculée concernant la situation au regard de l'activité économique peut être extrêmement utile pour l'établissement de tableaux des résultats, mais il exige des données provenant de différentes variables. Si l'on utilise les catégories définies dans la publication *Principes et recommandations concernant les recensements de la population et de l'habitation, deuxième révision*, une reconfiguration de plusieurs variables est nécessaire. La variable calculée pourrait comprendre deux catégories, elles-mêmes divisées en un ensemble de six sous-catégories :

1. Population active
  - 1.1 Personnes ayant un emploi
  - 1.2 Personnes au chômage
2. Population inactive
  - 2.1 Étudiant
  - 2.2 Personne s'occupant du foyer
  - 2.3 Retraité ou rentier
  - 2.4 Autres

35. Les différentes classifications au regard de l'activité économique étant utilisées dans nombre des travaux connexes, l'équipe de vérification doit examiner la possibilité d'introduire une variable calculée dans les enregistrements de données, plutôt que de prévoir un traitement comportant une reclassification de la situation concernant l'activité économique lors de l'établissement de tableaux. En effet, cette reclassification risque d'introduire des erreurs en ce sens que les différentes personnes chargées de définir le traitement des données risquent de mettre au point des séquences de reclassification légèrement différentes; ainsi, même un seul programme risque de reclasser différemment les informations en fonction des exigences particulières de la vérification ou de la mise en tableaux. Des spécialistes des classifications économiques doivent élaborer les spécifications concernant la variable calculée.

### 2. Propres enfants

36. Parfois, les pays veulent recueillir des informations sur les « propres enfants », qui sont les enfants biologiques du chef de ménage et/ou du conjoint. Les tableaux pourraient indiquer « propres enfants » qui seraient ensuite divisés en enfants vivant avec leurs deux parents ou seulement avec l'un de leurs parents.

37. La variable calculée pour les « propres enfants » pourrait être la somme du nombre des propres enfants d'une personne particulière, généralement du sexe féminin, selon les définitions choisies par les équipes de vérification. Parfois, les utilisateurs ont besoin d'informations plus détaillées sur les « propres enfants » par âge. Par exemple, aux États-Unis, des variables calculées sont établies pour le nombre de propres enfants de moins de 6 ans et pour ceux âgés de 6 à 17 ans. Ces valeurs sont inscrites sur les enregistrements de toutes les femmes. Ces informations sont utilisées spécialement pour déterminer les caractéristiques des femmes appartenant à la population active qui ont leurs propres enfants.

### 3. Parents vivant à la maison

38. Ces données présentent les caractéristiques des enfants de familles monoparentales par comparaison aux unités d'habitation dans lesquelles vivent les deux parents. Le programme de vérification établit cette variable calculée en déterminant le nombre de parents d'une personne donnée présents au foyer, sur la base des codes attribués aux liens de parenté. Le programme examine le code attribué aux liens de parenté dans le cas de chaque enfant et utilise conjointement cette information avec celle qui concerne les sous-familles pour déterminer le nombre de parents qui vivent dans l'unité d'habitation.

### 4. Scolarisation durant l'année en cours

39. Certains pays posent deux questions au sujet de l'éducation :

- a) La personne fréquente-t-elle l'école actuellement ?
- b) Quel est le niveau d'instruction atteint ?

40. Dans ces pays, les équipes de vérification constatent fréquemment un défaut de concordance entre les deux réponses lorsqu'une personne poursuit effectivement des études au moment du recensement. Parfois, le niveau le plus élevé atteint par la personne considérée risque d'être inférieur d'une année au niveau d'étude du moment. Si la personne se trouve à un niveau intermédiaire d'un cursus marqué par des niveaux successifs les statistiques n'en seront pas modifiées. Toutefois, si la personne en question suit la première année d'étude d'un cursus en vue de l'obtention d'un niveau déterminé, il ne sera sans doute pas possible d'établir une correspondance avec les données provenant d'autres sources. Par exemple, une personne inscrite en première année d'étude sera enregistrée comme étant scolarisée mais ne sera nullement censée avoir achevé un cycle d'étude quelconque. De même, quelqu'un qui commence des études secondaires sera enregistré comme étant scolarisé, mais son niveau d'instruction sera la dernière année d'études primaires.

41. On peut définir une variable calculée intitulée « niveau scolaire du moment » correspondant à cette combinaison de rubriques. Si la personne ne fréquente pas actuellement l'école, le code qui lui sera attribué sera le niveau de formation le plus élevé. Si la personne fréquente actuellement l'école, le programme de vérification ajoutera une unité au niveau de formation (ou au nombre d'années d'étude) et attribuera cette valeur à la variable « niveau scolaire du moment ».

42. Certains pays posent trois questions concernant l'éducation, à savoir les deux questions ci-dessus et une troisième sur le point de savoir si la dernière année d'étude a été achevée. Si cette information est également recueillie, elle devrait servir en outre à déterminer le « niveau scolaire du moment ».

### 5. Mois écoulés depuis la dernière naissance

43. Si des informations sur la question relative à la date de naissance sont recueillies, un nouveau code peut être créé pour établir des estimations indirectes de l'âge spécifique année par année et de la fécondité totale. Ce code prend la date du recensement, généralement le mois et l'année, puis détermine par soustraction le nombre de mois écoulés depuis la dernière naissance. Ce chiffre est consigné sur l'enregistrement de la femme pour aider à établir des estimations de la fécondité année par année.





## Annexe II

# Lien entre la présentation du questionnaire et l'introduction des données

1. Les deux formats de questionnaire les plus courants pour recueillir des informations dans le cadre des recensements ou des enquêtes sur la population sont les pages concernant les personnes et celles concernant les ménages.

2. La partie des questionnaires relative aux personnes se compose d'une page simple ou d'une double page où sont recueillies les informations démographiques, avec une page différente pour chaque personne. Cette méthode est utile parce qu'elle permet de faire figurer sur une même page toutes les informations concernant une personne, ce qui facilite la collecte. En outre, cette méthode permet de mieux contrôler la cohérence interne des données au cours du recensement. Les fiches personnelles peuvent être regroupées sous la forme d'un cahier pour en faciliter la manipulation sur le terrain, comme le montre la figure A.II.1.

3. Le codage et la saisie des données à inscrire sur les pages personnelles constitue essentiellement une opération mécanique, pour laquelle le codeur/l'opérateur de saisie n'est pas censé évaluer la validité des informations fournies mais plutôt leur at-

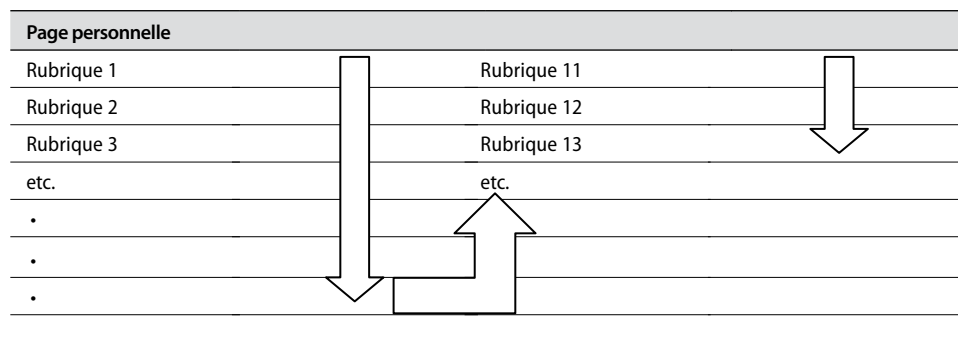
Figure A.II.1.

### Exemple de questionnaire constitué de pages personnelles

Page personnelle concernant la personne X		Page personnelle concernant la personne X + 1	
Rubrique 1	Rubrique 10	Rubrique 1	Rubrique 10
Rubrique 2	Rubrique 11	Rubrique 2	Rubrique 11
.	.	.	.
.	.	.	.
.	.	.	.

Figure A.II.2

### Exemple de séquence de collecte de l'information sur les pages personnelles d'un questionnaire



tribuer le code ou la touche de clavier appropriée. La figure A.II.2 décrit le flux d'informations concernant une personne donnée, enregistrées sur une page particulière. Il est plus facile d'inscrire les données sur une page unique concernant cette personne que de les introduire au clavier en tournant des pages. Les contrôles de validité sont effectués ultérieurement au cours des vérifications automatiques.

4. Les feuilles de recensement des ménages présentent l'ensemble des informations sur un ménage, si possible sur une seule page ou sur une série de pages, avec énumération de tous les membres du ménage sur chaque page. Cette façon de répertorier les membres du ménage est intéressante car les questions posées n'ont pas à être imprimées pour chaque individu, ce qui économise de la place. En outre, l'agent recenseur peut comparer les réponses fournies par les différents membres du ménage au fur et à mesure de leur collecte (figure A.II.3).

Figure A.II.3

**Exemple de questionnaire, page concernant un ménage, sur laquelle figurent toutes les personnes composant le ménage**

Page des ménages					
Personne	Rubrique 1	Rubrique 2	Rubrique 3	Rubrique 4	Etc.
1					
2					
3					
4					
5					
.					
.					
.					

5. Une troisième méthode consiste à prévoir une feuille distincte pour chaque personne, l'agent de recensement les regroupant en un cahier séparé pendant ou après le dénombrement. Cette méthode est efficace car l'agent de recensement réunit le nombre exact de feuilles (pages) nécessaires pour le ménage considéré. Son inconvénient tient au fait que les feuilles peuvent se détacher au cours des manipulations, ce qui peut poser des problèmes éventuels de vérification et de couverture si le bureau de recensement n'est pas en mesure de réunir les fiches correspondant au ménage approprié.

6. Le format proprement dit des pages du questionnaire mérite également examen, non seulement eu égard au dénombrement, mais aussi du point de vue de la saisie. Au cours du codage et de l'introduction des données, le document doit reposer à plat sur la surface de la table de travail, et les codeurs ou les opérateurs de saisie doivent pouvoir repérer et examiner aisément les informations inscrites sur le formulaire.

7. Lorsque toutes les informations figurent sur une seule page, le personnel peut introduire aisément le contenu des pages concernant les ménages, et cette opération sera évidemment plus rapide, vu que l'opérateur n'aura pas à tourner les pages. La figure A.II.4. présente la séquence de traitement des informations figurant sur une page concernant un ménage.

8. L'introduction de données sur la population ou l'habitation occupant plusieurs pages peut cependant soulever des difficultés. Afin de les résoudre, le bureau national de statistique est susceptible d'opter pour l'une des deux méthodes ci-dessous.

Figure A.II.4

**Exemple de séquence de traitement des informations inscrites sur des pages de questionnaire concernant les ménages, avec plusieurs personnes par page**

Page concernant un ménage				
Personnes	Rubrique 1	Rubrique 2	Rubrique 3	Etc.
1	→			↻
2	→			
3	→			
4				
5				
.				
.				
.				

9. Les données peuvent être introduites une personne à la fois. L'opérateur peut introduire la ligne d'informations concernant une personne sur la première page de la série, puis passer à la seconde et aux pages suivantes. Après avoir rempli les pages pour la première personne, l'opérateur de saisie revient à la première page concernant le ménage en question, puis introduit les données concernant la deuxième personne, puis la troisième, etc. Ce type de saisie se déroule directement tant que l'opérateur peut rester sur la même ligne de données tout au long de l'introduction. Bien que l'on puisse concevoir des programmes informatisés de vérification afin de démêler les données enregistrées, lorsque les réponses personnelles sont affectées par une erreur à la ligne correspondant à une autre personne, le programme proprement dit est très difficile à élaborer.

10. Les données peuvent être entrées une page à la fois. L'opérateur de saisie peut introduire une page entière de données avant de passer à la page suivante. Dans ce cas, il introduit toutes les informations devant figurer sur la première page, indépendamment du nombre de personnes concernées. Ensuite, il tourne la page et introduit les autres informations concernant l'ensemble des personnes. Les sauts de champ peuvent ou non être pris en compte, suivant le type de saisie (avec ou sans vérification informatisée). De toute façon, au cours de la vérification informatisée, les enregistrements des différents jeux de données introduites devront être regroupés et il faudra alors corriger toutes les erreurs éventuelles concernant les numéros affectés aux personnes.

11. Dans l'exemple ci-dessous (figure A.II.5), les données démographiques concernant le ménage ne soulèvent aucun problème particulier car les agents de recensement ont recueilli une réponse pour toutes les rubriques et toutes les personnes.

Figure A.II.5

**Exemple de page concernant un ménage à plusieurs personnes ne posant aucun problème de saisie**

Page des ménages				
Personnes	Lien avec le chef de ménage	Sexe	Âge	Etc.
1	Chef de ménage	M	40	
2	Conjoint	F	35	

Page des ménages				
Personnes	Lien avec le chef de ménage	Sexe	Âge	Etc.
3	Enfant	F	18	
4	Enfant	M	12	
5	Frère/Sœur	M	35	
6	Frère/Sœur du conjoint	F	30	
7	Frère/Sœur	M	5	
8	Frère/Sœur	F	3	
etc.				

12. Toutefois, une deuxième feuille concernant le même ménage (figure A.II.6) pourrait présenter certains problèmes d'introduction de données. Par exemple, si le pays convient d'indiquer la langue utilisée uniquement pour les personnes de 5 ans et plus, cette information sera laissée en blanc pour la huitième personne, âgée de 3 ans. L'opérateur de saisie devrait laisser la case en blanc en ce qui concerne cet enfant; sinon, la vérification automatique cherchera par la suite à la corriger.

13. De la même manière, d'autres rubriques devraient être laissées en blanc, telles que les personnes dont l'âge est inférieur à l'âge minimal légal pour travailler, les femmes n'ayant pas encore atteint l'âge de procréer et les informations sur la fécondité pour l'ensemble des hommes. Sur la figure A.II.6, l'opérateur de saisie risque d'introduire de façon erronée les données de la personne n° 6 concernant le nombre d'enfants (en l'occurrence, 4) dans la case de la personne n° 5. La vérification automatique supprimera alors les données sur la fécondité concernant les femmes tout en risquant de ne pas imputer la valeur exacte.

Figure A.II.6

**Exemple de page concernant un ménage à plusieurs personnes, susceptible de poser des problèmes de saisie**

Page 2 des ménages				
Personnes	Langue	Population active	Nombre d'enfants nés	Etc.
1	Langue 1	Oui		
2	Langue 1	Non	3	
3	Langue 1	Non	0	
4	Langue 1			
5	Langue 1	Oui		
6	Langue 1	Non	4	
7	Langue 1			
8				
etc.				

14. Dans nombre de cas, un pays est tenu d'utiliser le questionnaire concernant les ménages pour des raisons de coût ou de place. Toutefois, lorsque l'effectif de la population est réduit ou si le pays a les moyens d'assumer les frais supplémentaires, le questionnaire constitué de feuilles personnelles contient généralement moins d'erreurs de concordance, dues à des erreurs de saisie par comparaison aux questionnaires relatifs aux ménages.

## Annexe III

# Lecture optique ou saisie manuelle des données

1. De nombreux pays utilisent des scanners, lecteurs optiques de marques (OMR) ou équipements de reconnaissance optique de caractères (ROC). Chacune de ces solutions offre des avantages par rapport à la saisie manuelle, dans la mesure où leur mise en œuvre s'avère souple et rationnelle et ne comporte pas de coûts excessifs. Or, beaucoup de pays, même certains qui se sont engagés sur la voie de la lecture optique, n'ont pas nécessairement les moyens d'assumer les frais initiaux ou les frais d'entretien permanents pendant et après le recensement. Cela dit, nombre de pays utilisent souvent les scanners acquis pour le recensement à d'autres fins, notamment pour le dépouillement d'autres enquêtes et de documents administratifs tels que les fiches d'entrée et de sortie du territoire. Certains pays peuvent également recourir à la sous-traitance pour leurs opérations de lecture optique ou louer des scanners pour la durée nécessaire à la conduite du recensement.

2. L'introduction manuelle des données présente notamment l'avantage d'étendre le bénéfice des compétences acquises à cette occasion à d'autres activités menées dans le cadre des bureaux nationaux de recensement/statistique et d'autres administrations publiques. Lorsque le recensement a formé des opérateurs expérimentés de saisie des données, ces mêmes opérateurs sont ensuite en mesure d'effectuer cette tâche pour différentes enquêtes de suivi, notamment des enquêtes postdénombrement (PES) et différentes enquêtes concernant par exemple la fécondité ou encore les dépenses et les revenus des ménages. Ce personnel peut en outre assurer la saisie de documents administratifs divers (état civil, commerce, immigration et douane).

## A. Introduction des données

### 1. Lecture optique

3. Les pays qui utilisent des scanners ou d'autres dispositifs de lecture optique pour saisir leurs données ne les corrigent généralement pas au fur et à mesure, quoique les modifications éventuelles puissent dépendre des instructions de saut de champ intégrées au système. Le choix de la saisie manuelle des données comporte cependant différentes options suivant la rapidité d'introduction requise et l'importance des contrôles manuels nécessaires : chaque option est subordonnée des équipes de vérification, aux compétences des opérateurs et à la complexité du programme de vérification.

4. La quantité et le type de matériel nécessaire pour la saisie dépendent de la méthode choisie, des délais impartis pour cette phase des opérations, de la taille du pays, du degré de décentralisation de la saisie et de plusieurs autres facteurs. En saisie au clavier, la cadence moyenne de frappe va de 5 000 à 10 000 caractères à l'heure, cer-

tains opérateurs étant bien en dessous de cette moyenne, tandis que d'autres sont bien au-dessus. Les facteurs qui influent sur la cadence sont notamment : *a*) le programme et le logiciel choisis; *b*) la complexité des tâches confiées aux opérateurs; *c*) les caractéristiques ergonomiques, la fiabilité et la rapidité du matériel; *d*) l'arrivée régulière de travail; *e*) la formation et les compétences du personnel; et *f*) la diligence des opérateurs (Nations Unies, 2007, par. 1.193).

## 2. Saisie manuelle continue

5. L'introduction continue s'effectue sous deux formes. La première consiste à introduire toutes les données au fur et à mesure de leur apparition, sans sauter de champ. La saisie est alors plus rapide, puisque les opérateurs n'ont pas à s'arrêter quand ils trouvent des informations non valides ou incohérentes; elle peut en outre s'avérer plus précise, parce que plus mécanique. Le deuxième type d'introduction continue implique des interruptions pour vérifier la validité ou la cohérence des réponses indiquées dans le questionnaire, de telle sorte que les opérations se déroulent plus lentement et exigent beaucoup plus de compétences de la part du personnel. Il faut sérieusement tenir compte du prix élevé de cette option, eu égard à sa moins grande rapidité. Paradoxalement, cette méthode peut également améliorer la qualité lorsque les opérateurs constatent que les données ont été correctement enregistrées, mais mal codées. Quant aux erreurs d'introduction proprement dites, elles sont parfois immédiatement corrigées si le logiciel de vérification assure un contrôle automatique.

### *a*) Introduction continue sans sauts de champ

6. Lorsque toutes les données sont introduites ou sautées manuellement, il est possible de maintenir un certain rythme de travail et certaines instructions de saut de champ n'élimineront pas des données valides mais momentanément incohérentes. Par exemple, si une personne est déclarée comme étant de sexe masculin, les vérifications exigeront généralement de sauter toute la section de données concernant la fécondité. En l'occurrence, l'opérateur sautera les champs correspondants (au moyen de la barre d'espacement ou de la flèche pour sauter l'enregistrement correspondant à un homme ou à une femme trop jeune) tous les champs devant être laissés en blanc. Toutefois, cela prend du temps et l'espacement risque d'être inexact. Par exemple, l'opérateur peut introduire un espace excessif ou insuffisant, et les autres données sont alors mal introduites par suite d'une erreur d'alignement. Si tous les champs sont saisis de cette façon, alors ces informations peuvent être saisies sans aucune instruction de saut de champ. Ainsi, lorsqu'un opérateur doit introduire les informations concernant une femme adulte avec des données sur la fécondité (c'est-à-dire ayant indiqué un nombre d'enfants, un nombre d'enfants vivants ou un nombre d'enfants nés au cours de l'année écoulée, ces réponses ayant été codées), tous ces éléments sont introduits au clavier. Si les informations concernant la fécondité sont introduites au clavier, le programme de vérification informatisé peut déterminer quelles données ou séries de données sont valides et lesquelles d'entre elles doivent être modifiées. Si le programme de vérification établit que la personne considérée est une femme adulte, mais que les informations correspondantes de fécondité sont laissées en blanc, il faut alors obtenir les données nécessaires sur la fécondité pour la mise en tableaux des résultats en procédant par imputation dynamique ou par tout autre moyen approprié. En cas de perte de l'information effectivement recueillie en raison des instructions de saut de champ, l'équipe de vérification doit décider si le gain d'efficacité et de rapidité justifie cette perte. En présence d'instructions de saut de champ, les opérateurs peuvent néanmoins remonter aux écrans précédents pour retrouver l'emplacement des corrections introduites. S'il

est vrai qu'ils sont appelés à perdre du temps pour franchir les champs correspondant aux éléments non introduits, avec ce mode d'introduction des données, les incohérences entre sexe, âge et données sur la fécondité peuvent être résolues au cours des vérifications plutôt qu'au stade de la saisie.

#### b) Introduction continue avec sauts de champ

7. Une deuxième méthode d'introduction continue des données implique une saisie au clavier, avec respect des instructions de saut de champ. Là encore, si l'équipe de vérification exige l'observation de ces instructions, généralement pour reproduire le mode de collecte des données par les recenseurs, la saisie au clavier est facilitée et accélérée lorsque les instructions sont faciles à suivre et si elles sont assimilées rapidement par les opérateurs; si lesdites instructions sont particulièrement complexes, les opérateurs risquent d'être induits en erreur et d'introduire systématiquement les données au mauvais endroit. En présence de ce type d'instructions, l'efficacité maximale de saisie est obtenue avec des instructions limitées couvrant des parties importantes de l'enregistrement.

8. L'équipe de vérification devra définir les instructions de saut de champ adaptées à l'enquête ou au recensement entrepris dans le pays. Par exemple, il est logique de sauter toutes les rubriques sur l'emploi dans le cas des enfants, c'est-à-dire des personnes n'ayant pas encore atteint l'âge minimal pour travailler fixé dans le pays considéré. Souvent ces rubriques représentent la moitié des rubriques concernant la population, et il est donc parfaitement indiqué de les sauter dans le cas des enfants, sauf dans les situations spéciales comme celle des enfants d'âge limite ou si l'on veut enquêter sur le travail des enfants.

9. L'équipe de vérification examine les rubriques une par une afin de décider de celles qui seront prises en compte pour tel ou tel groupe d'âge. Les opérateurs peuvent regrouper les rubriques de façon à faciliter les sauts de champ.

10. Les décisions concernant les sauts de champs ne sont pas toujours faciles à prendre. Par exemple, si l'on considère la séquence ci-dessous :

1. Quelle est la nationalité de cette personne ?
  - Née dans le pays (passer à la rubrique 3)
  - Naturalisée
  - Non citoyenne
2. Quelle est l'année d'arrivée de cette personne ?
3. RUBRIQUE SUIVANTE

11. Une instruction de saut de champ pourrait être définie pour le passage de la rubrique 1 à la rubrique 3, c'est-à-dire pour sauter la rubrique concernant l'année d'arrivée pour les personnes nées dans le pays. Toutefois, les opérateurs de saisie ne respectent pas toujours les instructions de saut de champ, soit parce que l'agent recenseur ou le codeur a déjà commis une erreur, soit parce que les opérateurs en commettent une eux-mêmes. Parmi les nombreux facteurs en cause figurent le niveau de compétence des opérateurs, le contexte culturel, la disposition du questionnaire et la disposition des écrans de saisie. L'équipe de vérification cherche souvent à déterminer s'il est justifié de formuler des instructions de saut de champ dans un cas de ce type.

### 3. Saisie manuelle discontinue

12. Bien que l'introduction discontinue de données soit une méthode applicable dans le cadre d'un recensement, elle convient davantage aux enquêtes, en parti-

culier aux petites enquêtes, dont les résultats risquent d'être affectés par les réponses imputées. Selon les informations disponibles pour apporter des modifications ou introduire des corrections, la saisie discontinue peut impliquer des corrections manuelles ou automatisées.

13. Considérons le cas d'une petite enquête. Dans ce cas, chaque réponse est importante. Si un pays effectue une enquête sur un échantillon de 1 % de sa population, chaque réponse représente par exemple 100 personnes, ou 100 unités d'habitation ou 100 exploitations agricoles. Par conséquent, un petit nombre de réponses non valides ou incohérentes pourrait avoir une incidence considérable sur les résultats de l'enquête. En pareil cas, les démographes et autres spécialistes des sciences sociales préfèrent généralement avoir un sérieux droit de regard sur le traitement des données.

14. Ce contrôle peut s'exercer de différentes façons. Les démographes comme les autres spécialistes peuvent introduire les données eux-mêmes et vérifier au fur et à mesure l'absence de réponses aberrantes, non valides ou incohérentes, d'après les données enregistrées sur les formulaires de collecte. Ils sont souvent en mesure de résoudre immédiatement les incompatibilités, les erreurs de codage ou autres incohérences en examinant directement les informations recueillies. Ils peuvent parfois choisir de renvoyer sur le terrain des questionnaires incomplets ou non valides. Ce type de saisie discontinue donne de meilleurs résultats car le démographe fait alors en même temps fonction d'opérateur de saisie, mais il est de loin le plus coûteux, et nombre de pays n'ont pas les moyens d'y recourir.

15. Les équipes de vérification peuvent mettre au point des règles très précises pour définir ce que les opérateurs de saisie doivent faire en présence de cas particuliers. Pour chaque problème non résolu de code non valide, ils peuvent décider de ce que l'opérateur doit introduire. L'équipe de vérification peut résoudre les cas non couverts par les règles détaillées et modifier le cas échéant lesdites règles (mais au risque d'incohérences entre les données introduites avant l'établissement de ces règles et celles introduites après leur adoption).

16. Les instructions de saut de champ jouent un rôle important aussi bien dans le cas de l'introduction continue que dans celui d'une saisie discontinue. En effet, les opérateurs de saisie doivent être au courant et connaître parfaitement toutes les instructions de saut en vigueur. Comme on l'a vu précédemment, celles-ci peuvent accélérer l'introduction des données, mais généralement au prix d'une certaine perte de qualité. Dans le cas de la saisie discontinue, d'après une règle empirique généralement observée, la qualité est d'autant meilleure que les sauts de champ sont moins nombreux.

17. Une fois les instructions de saisie établies, les bureaux nationaux de recensement/statistique doivent demander aux opérateurs de saisie de tester ces instructions avant de décider des modalités d'exécution des tâches, notamment d'utiliser ou non la méthode de saisie dite interrompue. La mise à l'essai des instructions de saisie permet d'éliminer les erreurs du système et d'optimiser la procédure de saisie.

## **B. Contrôle**

18. Le bureau national de recensement/statistique doit également décider du niveau de contrôle approprié. Pour les données saisies, beaucoup d'experts recommandent un contrôle à 100 %. Dans ce cas, on réintroduit toutes les données (par-dessus les informations existantes) pour s'assurer que les données recueillies sont celles qui sont introduites dans la machine en vue de leur traitement informatique. Souvent, toute-



fois, la vérification intégrale n'est pas réalisable, soit parce que le pays ne dispose pas du temps nécessaire pour réintroduire toutes les données, soit parce qu'il ne dispose pas des ressources financières ou humaines requises. L'échantillon vérifié devrait être plus gros en pourcentage pour les opérateurs de saisie débutants mais plus faible pour les opérateurs expérimentés. En outre, si le test fait apparaître un taux d'erreur de saisie très faible, les erreurs imputables aux opérateurs étant très rares, un contrôle complet n'est probablement pas nécessaire.

19. Dans toute opération de contrôle, il importe d'abord de déterminer quelles sont les informations nécessaires. Le pays désire-t-il suivre chaque opérateur individuellement ou préfère-t-il les suivre par équipe ? Cherche-t-il à déterminer si l'opération permet l'acquisition de compétences ou leur maintien ? Les contrôles peuvent être importants, et prévoir des rapports journaliers, hebdomadaires ou mensuels, et ainsi de suite, pour déterminer les modalités d'exécution du travail et les compétences acquises.

20. Enfin, il est très important que le contrôle soit effectué en toute indépendance et qu'un autre ensemble d'opérateurs ou, du moins, différents membres de la même équipe, effectuent le contrôle depuis l'introduction des données. Le recours à un autre ensemble d'opérateurs confère plus d'indépendance au déroulement des opérations et, par conséquent, de meilleurs résultats.

21. Pour les données obtenues par lecture optique, il importe également d'effectuer un contrôle afin de s'assurer que la lecture a été détaillée et complète. Comme la technologie de la lecture optique est encore nouvelle, même lorsque les systèmes sont soigneusement testés au moyen de données pilotes ou prévérifiées, les variations de la qualité du papier et l'impression des formulaires en divers endroits, etc., peuvent être sources de problèmes qui devront être résolus lors du contrôle.

22. Si les erreurs sont systématiques et peuvent être supprimées par le programme de vérification, les opérateurs et les vérificateurs ne doivent pas porter de jugement sur la correction. Cependant, c'est aux opérateurs et aux vérificateurs qu'il appartient de découvrir les erreurs. Celles-ci peuvent être dues à un contrôle insuffisant de l'équipement de lecture optique, qui peut être source d'erreurs systématiques pour certaines rubriques ou combinaisons de rubriques, de confusion dans la lecture de certains chiffres (par exemple, en interchangeant 2s et 3s, ou 8s et 9s), d'erreurs de lecture des cases à cocher, et ainsi de suite.

23. La lecture erronée des cases à cocher est apparue comme un problème récurrent depuis quelques années, auquel il ne peut être que parfois remédié lors de la vérification. Si les formulaires ne sont pas contigus, il faut d'autres procédures, le plus vraisemblablement lors des vérifications structurelles, pour résoudre les problèmes. Comme on l'a indiqué précédemment, il importe de créer des fichiers structurés totalement sains avant d'entreprendre la vérification de leur contenu.

## 1. Contrôle par comparaison

24. Les techniques de vérification procèdent par comparaison de façon dépendante ou indépendante. Dans le premier cas, les opérateurs introduisent les données sur celles précédemment introduites par d'autres agents. Lorsque les caractères introduits au clavier sont différents, le progiciel en informe l'opérateur et, selon le programme, les données précédentes sont écrasées, ou bien la différence est notifiée. Comme les données sont introduites sur la base des questionnaires originaux, en règle générale, l'opérateur de saisie peut lui-même prendre une décision en connaissance de cause quant à l'exactitude éventuelle des données initialement introduites.

## **2. Contrôle indépendant**

25. Suivant la technique du contrôle indépendant, les opérateurs reprennent depuis le début l'introduction des données; ils créent un fichier entièrement indépendant sur la base des questionnaires originaux. Les deux fichiers ainsi obtenus, d'une part, le jeu de données initialement introduites et, d'autre part, le jeu de données de contrôle, sont alors comparés au moyen d'un programme informatique afin de vérifier les différences entre les deux. On peut supposer que la rectification des caractères non valides ou incohérents nécessite quelques interventions manuelles.

## **C. Considérations relatives à la vérification des données saisies par lecture optique**

26. De plus en plus de pays saisissent leurs données par lecture optique. Au début des années 2000, nombre de ces pays ont constaté avec surprise que la lecture optique introduisait des erreurs de types différents de ceux découlant de la saisie manuelle des données. Le problème lié à la vérification des données saisies par lecture optique tient en partie à l'absence de contrôle de qualité durant le processus de saisie. Comme la technologie était toute nouvelle à cette époque, beaucoup de bureaux de statistique n'avaient pas les connaissances ni les équipements nécessaires pour mettre au point un contrôle de qualité approprié de toutes les informations. Un grand nombre des pays qui se sont dotés de procédures appropriées de contrôle de qualité ont finalement renoncé à les étendre à toutes les rubriques, de sorte que certaines rubriques à la fin d'une question, en particulier celles sur la fécondité, ont produit des données non valides ou incohérentes.

27. Bien entendu, un grand nombre des incohérences constatées dans les données saisies au clavier apparaissent également dans les données obtenues par lecture optique. Comme les questionnaires à passer au scanner nécessitent des repères pour aider la machine à les lire, les rubriques sont souvent présentées de certaines façons qui peuvent causer des difficultés aux recenseurs et aux personnes interrogées lors de la collecte des données. Les difficultés liées à ces rubriques doivent être résolues de façon systématique. Lorsque ces rubriques sont étroitement liées à d'autres rubriques, comme la religion et l'appartenance ethnique, on peut utiliser les vérifications courantes décrites dans ce texte.

28. Toutefois, il faut faire preuve de prudence lorsque les rubriques nécessaires à la planification et à la définition de politiques risquent de poser des problèmes. Généralement, la rubrique correspondant au sexe ne pose pas de problèmes parce qu'elle n'offre que deux possibilités. Cependant, comme on l'a indiqué précédemment, si l'opérateur doit généralement se limiter à n'introduire qu'un 1 ou un 2 (ou un code pour « inconnu »), une valeur quelconque peut apparaître dans les colonnes concernant le sexe, sous la forme d'autres chiffres, de caractères alphabétiques ou d'autres caractères. Par conséquent, il faut ajouter un certain type de vérification à ce que l'on faisait auparavant pour les données introduites au clavier pour tenir compte de ces valeurs diverses.

29. Les codes relatifs au lien avec le chef de ménage illustrent bien ce problème. Si les codes appliqués à ce lien sont représentés par un seul chiffre, comme il est indiqué dans le texte, cela ne devrait pas normalement poser de problèmes. En revanche, si l'on utilise des codes à deux chiffres, cela peut poser un problème lors de la lecture optique si le premier chiffre est incorrectement codé ou relevé par le lecteur optique. Normalement, si l'on utilise des codes 1 à 12, l'opérateur devra se borner à n'introduire

que ces codes, et le programme d'introduction détectera l'introduction d'un code illégal. Avec la lecture optique, pratiquement n'importe quoi sera accepté (cela dit, les programmes de lecture optique peuvent être conçus pour détecter les codes incorrects). En pareil cas, les codes incorrects doivent être remplacés lors de la vérification, faute de quoi ils risquent de poser toutes sortes de problèmes au stade de la mise en tableaux.

30. L'âge peut être parfois source de difficultés, surtout si l'on utilise trois colonnes (pour tenir compte des personnes âgées de plus de 100 ans); par conséquent, une analyse chiffre par chiffre peut s'avérer nécessaire — ce qui implique que l'on examine séparément tous les chiffres des unités, des dizaines et des centaines — pour effectuer une bonne vérification. Une fois qu'il est établi que l'âge a été saisi correctement, on peut utiliser la vérification habituelle.

31. Toutefois, lorsque l'âge et la date de naissance ont été enregistrés, les informations trompeuses peuvent créer des difficultés si une rubrique l'emporte sur les autres. Généralement, les spécialistes du sujet préfèrent utiliser la date de naissance et la date du recensement ou de l'enquête pour produire (par soustraction) l'âge exact et le comparer à l'âge déclaré. S'il manque un ou plusieurs chiffres, il faut veiller à faire en sorte que tous les autres chiffres soient utilisés convenablement pour établir la meilleure estimation de l'âge calculé pour la comparaison. Lorsque la lecture optique ne capte pas un seul chiffre, par exemple, la vérification doit en tenir compte pour produire la meilleure estimation de ce qui aurait dû être capté. En général, on ne se heurte pas à ce type de problème avec la saisie manuelle.

32. Les rubriques pour lesquelles la lecture optique se heurtait aux plus graves difficultés au début des années 2000 avaient trait à la fécondité — à savoir le nombre d'enfants, le nombre d'enfants vivants et le nombre d'enfants nés au cours de l'année écoulée ou au fil des années. Par exemple, lorsque la valeur établie pour cette rubrique dans un pays donnée a été de 17, 18 ou 19 enfants du sexe féminin décédés, ces données, si elles étaient restées non vérifiées, auraient été inutiles pour les besoins de la planification.

33. Les données sur la mortalité peuvent également poser des problèmes lors de leur lecture optique. Par exemple, pour les données saisies manuellement, s'il y a une série de rubriques concernant les décès survenus au cours de l'année qui a précédé le recensement (sexe et âge des personnes décédées, mort naturelle et mortalité maternelle), la saisie manuelle tient compte des inscriptions effacées ou rayées. En revanche, avec les données saisies par lecture optique, les inscriptions effacées ne seraient normalement pas lues, et le lecteur laisserait un blanc et poursuivrait la saisie. Le programme de vérification doit transférer les informations dans les espaces appropriés pour l'établissement des tableaux, puis pour l'analyse. Il convient de noter que les nouvelles opérations de lecture optique peuvent effectuer ces transferts pendant et juste après la saisie.

## D. Conclusions

34. Comme, malheureusement, les problèmes qui se posent à un pays dépendent de la programmation particulière et du fonctionnement des différents lecteurs optiques, il devient difficile de définir des directives générales. Cependant, dans tous les cas examinés jusqu'ici, les problèmes de lecture optique rencontrés ont été systématiques; autrement dit, une fois que le personnel a défini l'algorithme nécessaire pour faire face au problème, des jeux de données pleinement vérifiés ont pu être établis.



## Annexe IV

# Exemples d'organigrammes

1. L'une des tâches de l'équipe de vérification consiste à structurer les relations entre les variables utilisées lors du processus de vérification. La mise au point d'organigrammes facilite l'identification des différents liens entre les variables, ainsi que l'élaboration de spécifications de vérification claires et précises. Ces spécifications quant au lien entre les variables aident aussi bien les spécialistes du sujet que les informaticiens à visualiser le processus de vérification et facilitent les échanges entre les deux groupes.

2. On trouvera ci-dessous trois exemples d'organigrammes :

- a) Organigramme de détermination du chef de ménage;
- b) Organigramme de détermination de la présence d'un conjoint dans le ménage;
- c) Organigramme de vérification de la variable « sexe » du chef de ménage et de son conjoint.

Ces organigrammes sont présentés uniquement à titre d'exemples et doivent être considérés comme tels. L'équipe de vérification peut leur apporter les modifications nécessaires en fonction du contexte national.

3. Chacune des variables prises en compte dans un recensement doit faire l'objet d'un organigramme de vérification. L'équipe de vérification doit s'employer à mettre au point les organigrammes et les informaticiens doivent les utiliser avec les spécifications pour élaborer les programmes de vérification des données de recensement. Organigrammes et spécifications doivent faire l'objet d'une documentation adéquate pour pouvoir être utilisés lors du traitement de données des enquêtes et recensements futurs.

Figure A.IV.1  
Exemple d'organigramme de détermination du chef de ménage

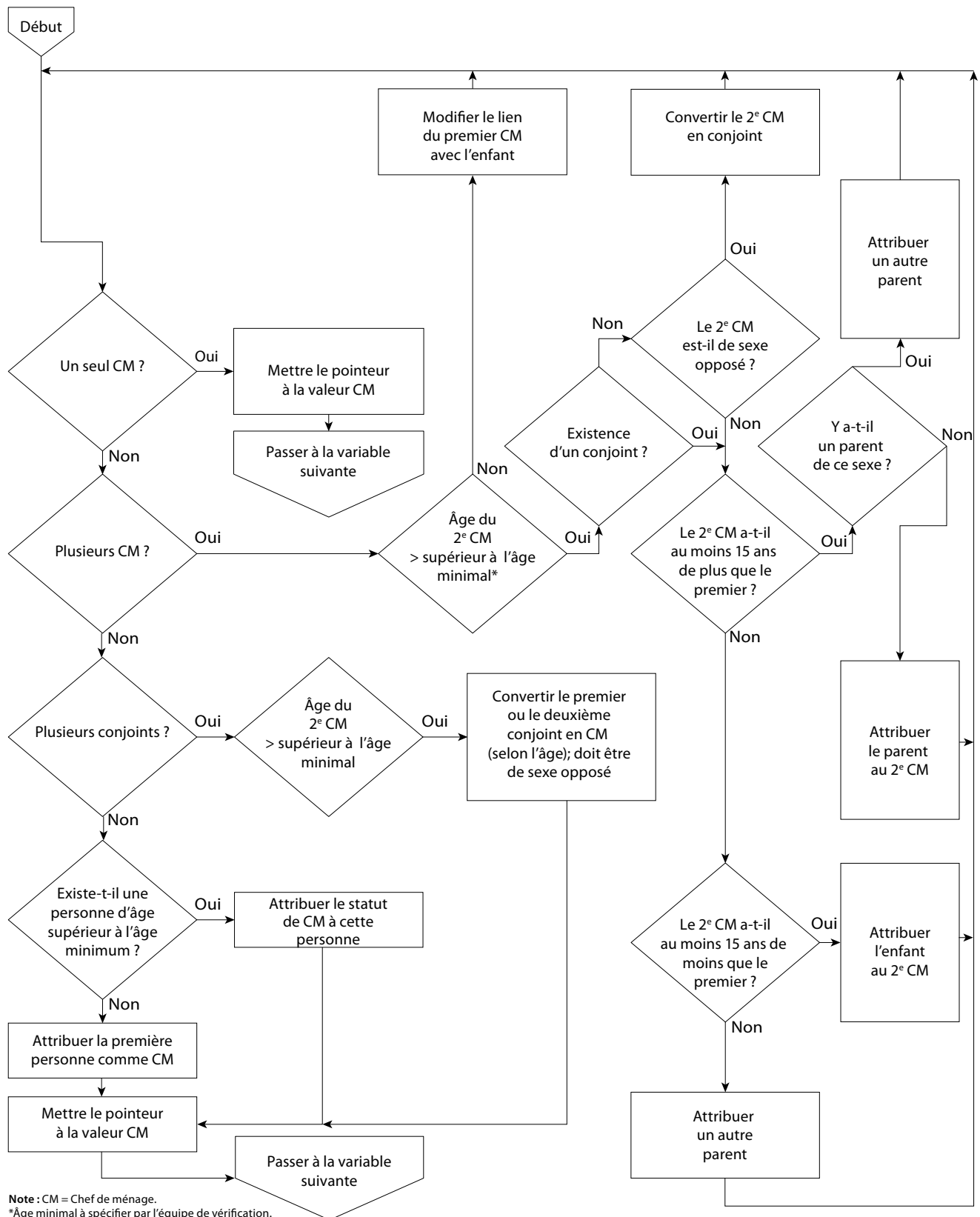
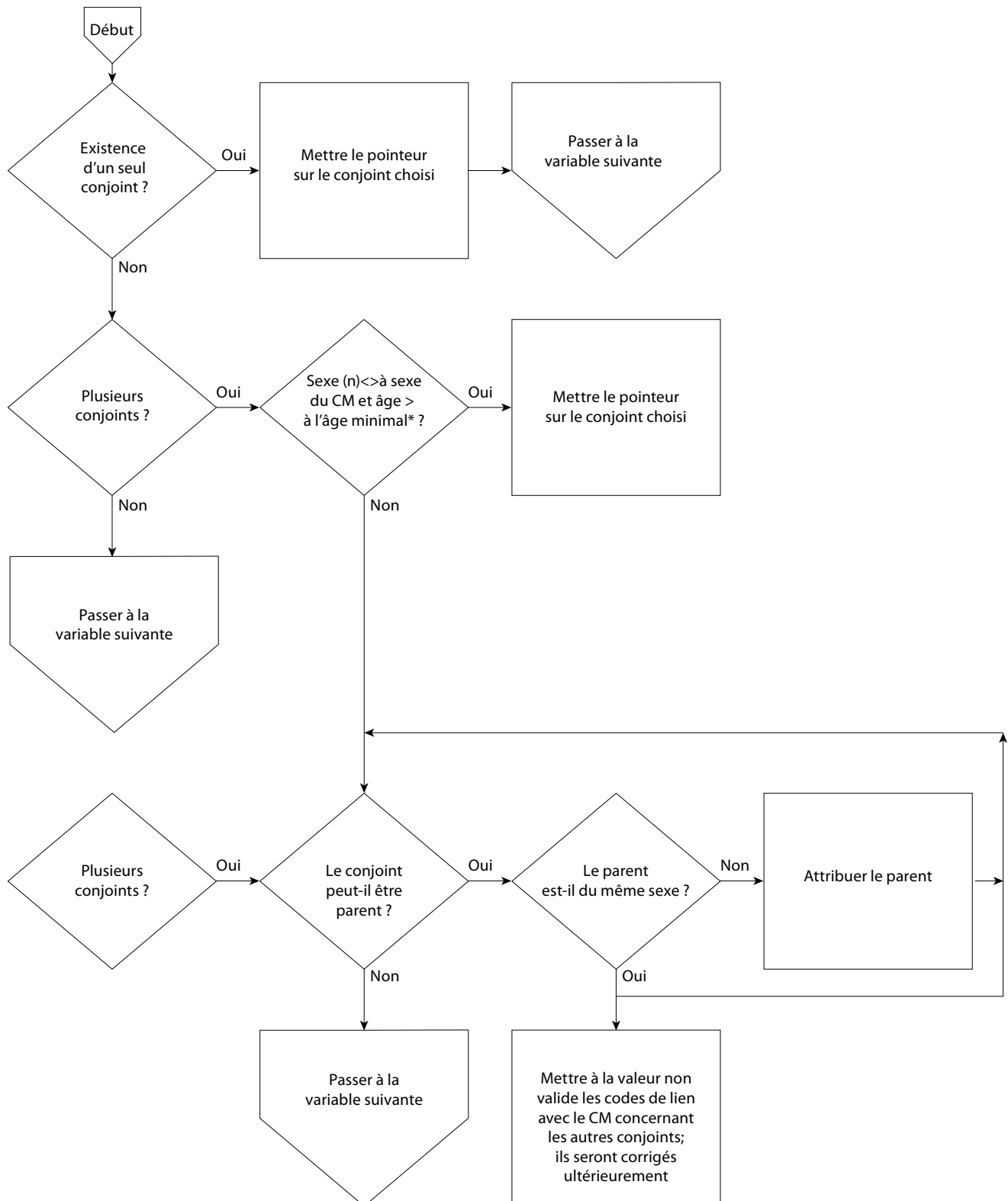


Figure A.IV.2  
Exemple d'organigramme de détermination de la présence d'un conjoint dans le ménage

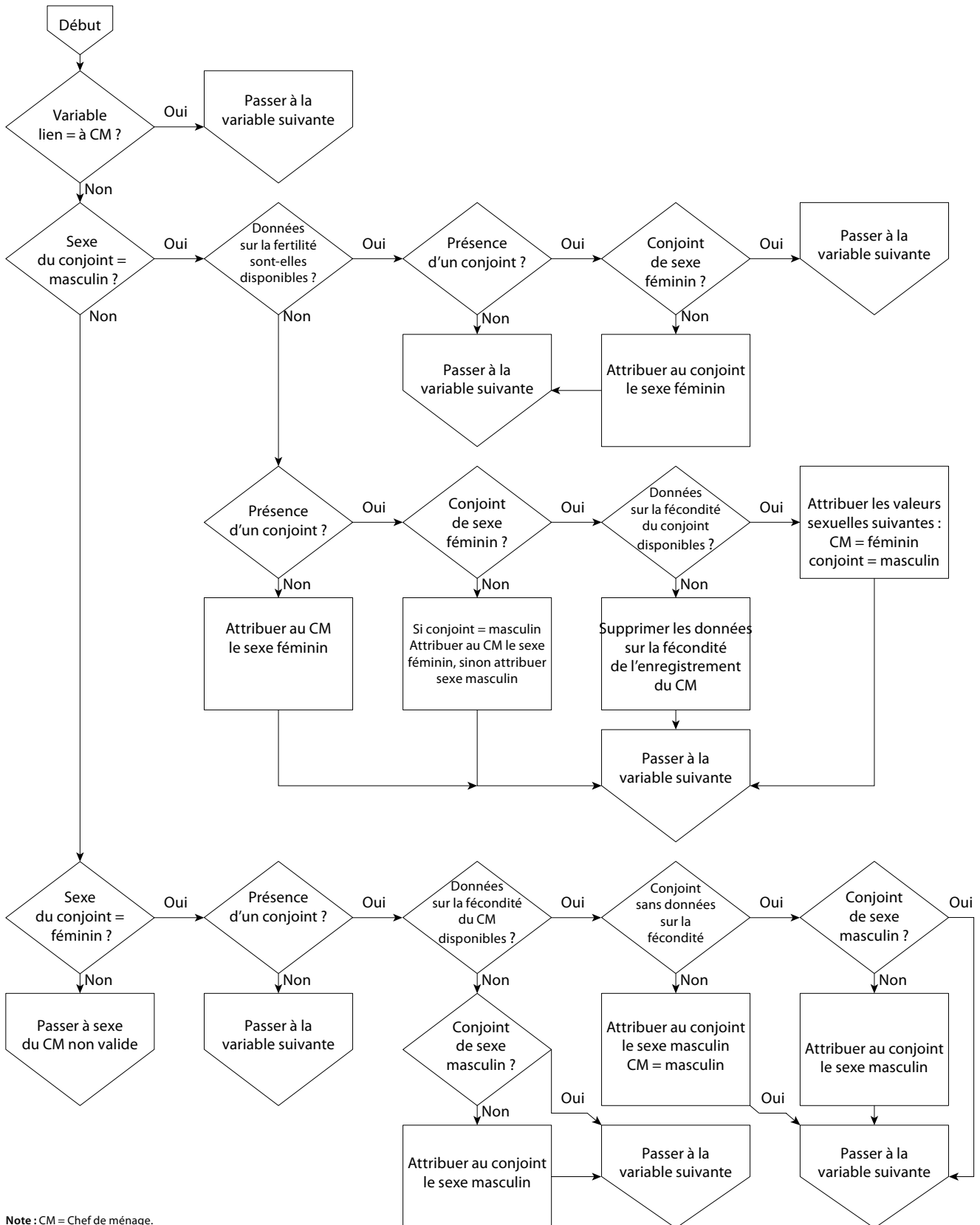


Note : CM = Chef de ménage.

\* Âge minimal à spécifier par l'équipe de vérification.

Figure A.IV.3

Exemple d'organigramme de vérification de la variable sexe concernant le chef de ménage et le conjoint

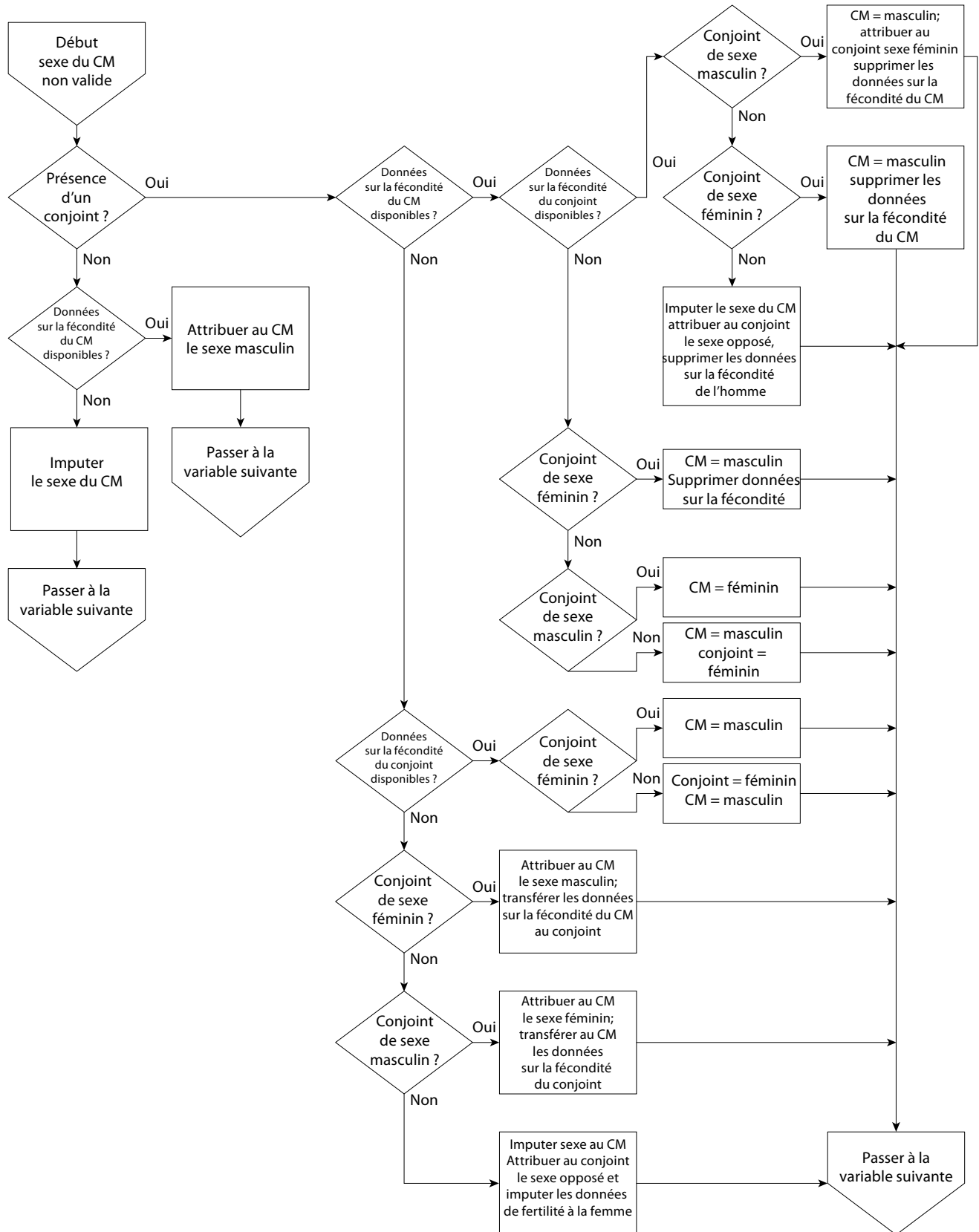


Note : CM = Chef de ménage.



Figure A.IV.3 (suite)

Exemple d'organigramme de vérification de la variable sexe concernant le chef de ménage et le conjoint





## Annexe V

# Méthodes d'imputation

1. Un certain nombre de méthodes d'imputation ont été mises au point. La plupart des méthodes décrites ci-dessous sont analysées dans les publications de Kalton et Kasprzyk (1982, 1986); Sande (1982); et Särndal, Swensson et Wretman (1992).

2. Les méthodes d'imputation peuvent être qualifiées de stochastiques ou de déterministes en fonction du caractère aléatoire des données imputées.

3. **Les méthodes d'imputation déterministes** comprennent l'imputation déductive, les méthodes d'imputation par modélisation, par exemple, par la moyenne et par régression et (le cas échéant) par voisin le plus proche.

4. **L'imputation déductive** est une méthode qui permet de déduire avec certitude une donnée dont la valeur est manquante ou incohérente. Ce type d'imputation sera souvent réalisé en fonction de l'ensemble des réponses fournies au titre des autres rubriques du questionnaire.

5. Plus généralement, la technique d'imputation doit remplacer la valeur entachée d'incertitude par la valeur exacte. Certaines des procédures d'imputation d'usage courant sont présentées dans les paragraphes suivants.

6. À l'exception des algorithmes d'imputation dynamique à partir d'un donneur unique, les méthodes décrites ci-après concernent les procédures d'imputation rubrique par rubrique. Ainsi, à l'intérieur de chaque classe d'imputation, les rubriques de l'enregistrement sont considérées l'une après l'autre, de manière séquentielle. En général, cela consiste à considérer uniquement les vérifications concernant explicitement la rubrique en question ou un ensemble restreint de variables étroitement liées. Compte tenu de l'existence possible de vérifications explicites ou implicites associant la (les) rubrique(s) en question à d'autres rubriques devant être examinées ultérieurement au cours du processus, une valeur imputée peut éventuellement (tout en satisfaisant aux vérifications en cours), entraîner des rejets à l'occasion d'autres vérifications ultérieures. Les valeurs imputées ne pourront de façon certaine franchir avec succès la totalité des vérifications que si l'on envisage la réalisation d'une série complète de vérifications, y compris de l'ensemble des vérifications implicites. On entend par vérification implicite une vérification pouvant être déduite par combinaison logique d'au moins deux vérifications explicites.

7. Dans les descriptions ci-dessous, l'expression « enregistrements acceptés » désigne les enregistrements qui ont subi avec succès toutes les vérifications concernant la (les) rubrique(s) en question. Les « enregistrements rejetés » désignent ceux qui ont été rejetés après au moins une vérification concernant la ou les rubriques en question.

8. **L'imputation de la moyenne générale** a pour effet d'attribuer la valeur moyenne de la rubrique (calculée pour les enregistrements acceptés) à la réponse manquante ou incohérente de tous les enregistrements rejetés. Cette méthode est susceptible de produire des estimations ponctuelles fondées, mais s'avère moins intéressante lorsque le calcul de la variance doit utiliser un estimateur standard. Les estimations de variance risquent d'être fortement sous-estimées sauf lorsque le taux d'imputation est

très faible ou en cas d'utilisation d'un estimateur de variance modifié tenant compte de l'imputation.

9. **L'imputation de la moyenne de classe** utilise les classes d'imputation définies pour créer des groupes d'enregistrements présentant un certain degré de similarité. À l'intérieur de chaque classe, la moyenne de la rubrique relative aux enregistrements acceptés est imputée aux réponses manquantes ou incohérentes de chacun des enregistrements rejetés. Bien que cette technique soit très proche de celle de l'imputation de la moyenne générale, l'incidence sur la distribution et les problèmes liés à l'estimation de la variance sont alors généralement nettement moins prononcés.

10. **L'imputation par régression** ou d'une façon plus générale **par modèle** utilise les données des enregistrements acceptés pour calculer selon une loi de régression la variable devant faire l'objet d'une imputation d'après une série de variables prédictives. Les variables prédictives de la régression peuvent figurer parmi les rubriques du questionnaire ou peuvent être des variables auxiliaires. La loi de régression permet ensuite d'imputer les valeurs manquantes ou incohérentes. Il s'agit d'un cas particulier de la technique d'imputation par modélisation. Cette méthode est souvent utilisée pour les variables continues des enquêtes économiques, caractérisées par la possibilité souvent observée de prévoir de façon satisfaisante les valeurs occasionnelles du moment d'après des valeurs occasionnelles antérieures.

11. **La méthode d'imputation par voisin le plus proche** ou **d'appariement par une fonction de distance** a pour effet d'attribuer une valeur rejetée par une vérification d'après l'enregistrement accepté « le plus proche », cette caractéristique de proximité étant définie au moyen d'une fonction de distance sur la base d'autres variables connues. Il s'agit d'une méthode applicable à l'intérieur de classes d'imputation. Elle est généralement considérée comme adaptée aux variables continues, mais peut également être appliquée aux variables non numériques.

12. **Les méthodes d'imputation stochastiques** comprennent les méthodes fondées sur une loi de régression ou toute autre méthode déterministe, avec addition d'éléments résiduels aléatoires, ainsi que les méthodes dites d'imputation dynamique (*hot deck*) ou statique (*cold deck*).

13. À chaque méthode déterministe correspond une méthode stochastique. Il suffit pour cela d'ajouter à la valeur obtenue par imputation déterministe un élément résiduel aléatoire provenant d'une distribution appropriée. Cette façon de procéder contribuera à mieux conserver la distribution de fréquences des données du fichier. Kalton et Kasprzyk (1986) passent en revue différentes versions de cette technique.

14. **Les méthodes d'imputation dynamique et statique [dites à « donne » variable (*hot deck*) et à « donne » fixe (*cold deck*)]** s'emploient à créer une variabilité plus réaliste des valeurs imputées par comparaison aux résultats obtenus avec les méthodes déterministes. Les méthodes d'imputation dynamique remplacent les valeurs manquantes ou incohérentes par des valeurs sélectionnées (de façon aléatoire) dans les enregistrements de l'enquête ou du recensement en cours qui ont subi avec succès les vérifications. Quant aux méthodes d'imputation statique, elles imputent des valeurs obtenues à partir de sources différentes, souvent des données historiques recueillies dans le cadre d'enquêtes ou de recensements antérieurs du même type. Il existe différentes formes d'imputation dynamique ou statique.

15. **L'imputation globale aléatoire** est la forme la plus simple d'imputation dynamique (*hot deck*). Pour chaque enregistrement rejeté, un enregistrement accepté est choisi au hasard dans l'ensemble des enregistrements acceptés et la valeur indiquée correspondant à la rubrique en question est imputée à l'enregistrement rejeté.

16. **L'imputation aléatoire à l'intérieur de classes** utilise également des classes d'imputation afin de limiter la sélection aléatoire de l'enregistrement donneur à un

ensemble censé présenter des similarités par rapport à l'enregistrement devant faire l'objet d'une imputation.

17. **La technique d'imputation dynamique séquentielle** fait également appel aux classes d'imputation et présente l'avantage de n'exiger qu'une seule passe dans le fichier de données pour achever le processus d'imputation. La procédure commence par l'attribution statique d'une valeur à chaque classe d'imputation, les enregistrements du fichier de données étant examinés successivement. Lorsqu'un enregistrement vérifié accepté est détecté, sa valeur pour la rubrique en question remplace la valeur enregistrée pour la classe d'imputation. Lorsqu'un enregistrement rejeté par la vérification est détecté, sa valeur manquante ou incohérente est remplacée par la valeur enregistrée. Le nombre de classes d'imputation ne peut être très élevé, puisqu'il faut être assuré de la disponibilité des donneurs dans chaque classe d'imputation. Si l'ordre d'enregistrement dans le fichier de données est aléatoire, cette méthode sera sensiblement équivalente à une imputation aléatoire à l'intérieur de classes. Elle a par ailleurs pour inconvénient d'impliquer dans nombre de cas une utilisation répétée des donneurs, ce qui risque d'affecter la distribution de la variable considérée et les estimations de variance.

18. **L'imputation dynamique hiérarchique** représente une amélioration de l'imputation dynamique séquentielle, qui consiste à utiliser un nombre élevé de classes d'imputation. Lorsqu'on ne peut identifier un donneur dans la classe d'imputation initiale, on regroupe les classes en ordre hiérarchique jusqu'à ce qu'un donneur soit trouvé.

19. Les **algorithmes d'imputation dynamique à donneur unique** ont pour objectif d'imputer des données à un enregistrement rejeté à partir d'un donneur unique. Ils permettent donc l'imputation conjointe de toutes les données d'un enregistrement identifié comme problématique par les vérifications. En pratique, l'objectif consiste souvent à utiliser un donneur unique pour chaque ensemble de variables étroitement liées de l'enregistrement. Cette approche présente l'avantage notable d'améliorer non seulement les distributions marginales à l'instar des méthodes d'imputation dynamique décrites ci-dessus, mais également les distributions de fréquences conjointes. L'autre avantage des méthodes d'imputation dynamique à donneur unique est qu'elles limitent le problème des valeurs d'imputation qui seront rejetées lors des vérifications envisagées pour des groupes de variables ultérieurs. Dans le contexte des méthodes d'imputation dynamique à donneur unique, on entend par « enregistrement accepté » un enregistrement qui a subi avec succès toutes les vérifications appliquées au groupe de variables. Par contre, un « enregistrement rejeté » est un enregistrement qui a été rejeté lors d'au moins une de ces vérifications.

20. **La méthode de vérification et d'imputation de Fellegi-Holt** (Fellegi et Holt, 1976) considère simultanément toutes les vérifications. Elle se caractérise essentiellement par le fait que les règles d'imputation sont tirées des vérifications correspondantes sans spécification explicite. Pour chaque enregistrement rejeté, la méthode prévoit initialement une phase de localisation de l'erreur au cours de laquelle elle détermine l'ensemble minimal de variables à imputer, et procède ensuite à l'imputation proprement dite. Dans la plupart des cas, on choisit un donneur unique parmi les enregistrements acceptés en procédant par appariement, d'après les autres variables impliquées par les vérifications mais ne devant pas faire l'objet d'une imputation. La méthode cherche à établir une correspondance exacte unique, et peut être étendue de façon à tenir compte des autres variables non explicitement liées aux vérifications. Parfois, aucun donneur adéquat ne peut être identifié, ce qui exige alors le recours à une méthode d'imputation par défaut.

21. **La méthode d'imputation par voisin le plus proche ou nouvelle méthode d'imputation (NMI)** [Bankier *et al.*, 1996; Bankier, Lachance et Poirier, 1999] s'apparente à la méthode de Fellegi-Holt en ce qu'elle considère simultanément toutes les vérifications, ne spécifie pas explicitement les opérations d'imputation et effectue les imputations en partant d'un donneur unique. Pour chaque enregistrement rejeté, elle identifie les imputations impliquant le moins de changements possibles, sous réserve de la disponibilité de donneurs potentiels. Cela garantit la disponibilité d'un donneur. Contrairement à la méthode de Fellegi-Holt, cette méthode cherche d'abord des donneurs, puis détermine les imputations nécessitant le moins de changements. Elle recherche des donneurs en procédant par appariement, utilisant toutes les variables (y compris celles qui pourraient être imputées) nécessaires aux vérifications, et peut se contenter d'appariements approximatifs pour les variables numériques complétés par des appariements exacts pour la plupart des autres variables, mais pas nécessairement pour toutes. Les imputations établies sur la base de chaque donneur potentiel sont déterminées, et celles qui impliquent le minimum possible de modifications sont identifiées. Cette méthode prend également en considération les imputations qui se rapprochent de celles qui impliquent un minimum de modifications, car celles-ci peuvent parfois produire des enregistrements imputés plus plausibles. Enfin, on choisit au hasard l'une des imputations impliquant le minimum de modifications et l'une de celles se rapprochant des précédentes, et l'on procède à l'imputation.

22. Bien que ces deux méthodes (Fellegi-Holt et NMI) exigent beaucoup de calculs, il existe des algorithmes efficaces qui permettent leur mise en œuvre sur des ordinateurs modernes. Tel est particulièrement le cas pour la nouvelle méthodologie d'imputation, qui peut traiter des problèmes de vérification et d'imputation sensiblement plus lourds que la méthode Fellegi-Holt.

23. Toutes les méthodes d'imputation ci-dessus permettent d'obtenir une valeur imputée unique pour chaque valeur manquante ou incohérente. Elles introduisent toutes une certaine distorsion dans la distribution habituelle des variables concernées et risquent de déboucher sur des estimations de variance inadéquates en cas d'utilisation d'estimateurs standard. L'ampleur de la distorsion introduite dépend pour beaucoup de la quantité d'imputations effectuées et de la méthode employée.

24. **L'imputation multiple** est une méthode proposée par Rubin (1987) qui implique plusieurs imputations ( $m$ ) pour chaque valeur à imputer. Ensuite, à partir du jeu complet de données, on peut obtenir  $m$  estimations de la variable en cause. À partir de ces estimations, on établit une estimation combinée unique, ainsi qu'une estimation de la variance combinée illustrant le degré d'incertitude quant au choix de la valeur à imputer. La méthode d'imputation multiple présente l'inconvénient d'exiger un surcroît de travail de traitement des données et de calcul d'estimations.

25. La plupart des systèmes d'imputation font appel à un ensemble de méthodes d'imputation; généralement, l'imputation déductive est utilisée quand cela est possible avant de recourir à une ou plusieurs autres procédures. La plupart des bureaux nationaux de statistique emploient une méthode ou une autre d'imputation dynamique pour les vérifications et les imputations des données de recensement. La méthode d'imputation dynamique séquentielle et celle de Fellegi-Holt sont celles dont l'usage est le plus courant. Parmi les bureaux nationaux de statistique qui utilisent actuellement la méthode de Fellegi-Holt, l'un d'entre eux lui substitue à présent la nouvelle méthode d'imputation (NMI) et plusieurs autres envisagent d'en faire autant. Toutefois, compte tenu de l'éventail escompté des lecteurs du présent manuel, cet ouvrage se concentre plus particulièrement sur une forme d'imputation dynamique séquentielle.

## Annexe VI

# Progiciels de vérification informatisée

1. Grâce aux micro-ordinateurs relativement bon marché actuellement disponibles, les pays devraient être en mesure de vérifier les données de leurs recensements et de leurs enquêtes en procédant de façon minutieuse et en temps voulu<sup>a</sup>. Récemment encore, chaque pays devait rédiger son propre programme de vérification selon ses besoins, ce qui comportait un coûteux investissement en temps, aussi bien pour l'élimination des erreurs que pour le traitement. Avec les nouveaux progiciels de vérification normalisés, il est aujourd'hui beaucoup plus facile de répondre aux besoins d'un pays en matière de vérification, et avec moins de compétences en informatique.

2. L'utilisation d'un progiciel de vérification a pour avantage que, s'il est convenablement utilisé, les données obtenues seront cohérentes et exemptes d'erreurs, de sorte que les tableaux pourront être produits en temps voulu. Nombre de progiciels, par exemple les systèmes d'analyse statistique (SAS) et les progiciels SPSS de statistiques pour les sciences sociales, ou divers langages perfectionnés, peuvent servir à l'élaboration de programmes de vérification. À défaut, le pays peut utiliser l'un des progiciels conçus expressément pour la vérification des données de recensements et d'enquêtes. Pour la plupart des pays, la vérification au moyen d'un progiciel d'usage général est beaucoup plus rapide qu'avec un programme spécialement conçu, car le premier n'exige pas le même niveau de connaissances informatiques que le second.

3. Un bon progiciel de vérification assure la communication entre les spécialistes du sujet et les programmeurs. Il devrait permettre d'introduire des commentaires ou des pseudocodes à côté des codes de programmation, à moins que ces codes ne soient accessibles et compréhensibles pour les spécialistes du sujet. Les démographes et les autres spécialistes devraient pouvoir consulter le programme ligne par ligne et comprendre exactement ce qu'il fait.

4. Tout progiciel de vérification existant qu'un pays envisage d'utiliser doit pouvoir fonctionner et produire des comptes rendus relatifs aux différents contrôles, tests et imputations qu'exige la vérification des données d'un recensement. Ces exigences ont cours même lorsque les informaticiens élaborent des programmes de vérification sur mesure. Le progiciel doit répondre aux exigences suivantes :

- a) Capacité de saisie et/ou de vérification des données introduites. Ce progiciel doit permettre l'introduction d'instructions de saut de champ. Par exemple, l'équipe de vérification peut décider de sauter les questions sur la fécondité lorsqu'elle interroge des hommes;
- b) Réalisation de vérifications structurelles, ce qui permettra d'établir la présence effective des types d'enregistrements requis, par exemple, d'un enregistrement concernant l'habitation pour chaque numéro séquentiel;
- c) Création d'enregistrements pour remplacer ceux qui manquent et/ou attribution de pondérations aux enregistrements existants;
- d) Confirmation que la valeur attribuée à chaque variable est valide;

a Il importe de souligner que le présent manuel se concentre sur la vérification des données de recensements. Les enquêtes étant limitées à un moins grand nombre de personnes à interroger, et comportant généralement plus de questions, elles nécessitent souvent la conception de programmes de vérification plus détaillés. Certains des programmes examinés dans la présente annexe sont conçus pour l'exécution d'enquêtes dans des pays à forte population. Cela dit, ces programmes conviendraient également pour les recensements effectués dans de petits pays. Lorsque les fichiers s'agrandissent, ils utilisent moins facilement certaines méthodes statistiques, telles que les analyses de régression et les analyses multivariées. En revanche, avec un fichier détaillé, l'impact de l'absence d'une réponse est moins sensible que dans une enquête où chaque individu sélectionné pour l'échantillon « compte » en fait pour beaucoup de personnes. C'est pourquoi chaque bureau national doit tester divers programmes afin de déterminer celui qui répond le mieux à ses besoins.

- e) Mise en mémoire de la totalité ou d'une partie des enregistrements déjà vérifiés;
- f) Contrôle de la cohérence entre plusieurs caractéristiques d'un même enregistrement ou d'un enregistrement à un autre. Un sous-ensemble de ces contrôles consiste à s'assurer de la cohérence au sein des ménages en comparant les réponses reçues à celles des membres du ménage précédent. Imputation de valeurs par la technique d'imputation dynamique (*hot deck*) si le pays choisit d'utiliser cette méthode;
- g) Utilisation de plusieurs valeurs tirées d'un enregistrement ou de multiples enregistrements pour établir la valeur d'une variable calculée, et introduction de celle-ci dans l'enregistrement approprié;
- h) Identification et élimination des enregistrements en double;
- i) Établissement d'un journal dans lequel sont consignées les erreurs et les modifications par petites zones géographiques.

5. Généralement, les progiciels ou les programmes vérifient un enregistrement à la fois; toutefois, certains progiciels actuels permettent d'effectuer des contrôles d'un enregistrement à l'autre, notamment pour les unités d'habitation.

6. Comme on l'a indiqué précédemment, jusqu'à l'introduction de la méthode de Fellegi-Holt (1976) et de ses versions plus récentes, presque toutes les vérifications utilisaient l'approche descendante, ce qui veut dire que les rubriques étaient vérifiées généralement mais pas toujours dans l'ordre dans lequel les informations avaient été recueillies. Par exemple, comme la première rubrique à vérifier concernant la population est généralement le « lien avec le chef de ménage », la rubrique « sexe » serait vérifiée sur la base de cette première rubrique, puis « l'âge » sur la base à la fois du sexe et du lien avec le chef de ménage, et ainsi de suite.

7. Au cours des dernières décennies, plusieurs systèmes d'imputation à minimum de modifications fondés sur le système de Fellegi-Holt ont été mis au point. Parmi ceux-ci, on peut citer CANEDIT et le Système généralisé de vérification et d'imputation (SGVI) de Statistique Canada, suivis du Système canadien de contrôle et d'imputation du recensement (SCANCIR) [Bankier, 2005; Chen, 2007<sup>b</sup>]. Les États-Unis d'Amérique ont conçu le système DISCRETE (Chen *et al.*, 2000; Winkler, 1997a; Winkler, 1997b; Winkler et Chen, 2002) et Structured Programs for Economic Editing and Referrals (SPEER) [Greenberg et Surdi, 1984; Winkler et Draper, 1997] au U.S. Census Bureau. Kovar et Winkler (2000) ont fait une comparaison assez détaillée des systèmes canadien et américain. Parmi les autres systèmes de vérification découlant du système de Fellegi-Holt, on peut citer le système de vérification CHERRYPI fondé sur le système de Fellegi-Holt conçu par Statistics Netherlands (De Waal et Van de Pol, 1997).

8. Le recensement effectué par le Canada en 1996 (et d'autres) utilisait une approche différente, appelée Nouvelle méthode d'imputation (NMI) par voisin le plus proche. La version 1996 imputait simultanément des réponses pour l'âge, le sexe, l'état matrimonial et le lien avec le chef de ménage pour toutes les personnes d'une unité d'habitation (Bankier, 1999). Cette méthode a été améliorée et élargie pour le recensement canadien de 2001 et pour les activités statistiques ultérieures (Bankier, Lachance et Poirier, 2000; Bankier, 2001).

9. La méthode d'imputation par voisin le plus proche recherche les donneurs les plus proches puis détermine les imputations minimales sur la base de ces donneurs. Alors que la méthode de Fellegi-Holt implique l'imputation du plus petit nombre de variables et la préservation de l'intégrité des sous-ensembles de population, la NMI,

**b** Dans un examen du SCANCIR, Bankier, Lachance et Poirier (2000, p. 10) résumant comme suit : le SCANCIR, avec ses algorithmes de vérification et d'imputation d'une grande efficacité, laisse entrevoir la possibilité de résoudre des problèmes d'imputation d'ordre très général comportant un grand nombre de règles de vérification et un grand nombre de variables qualitatives et quantitatives lorsqu'une imputation d'un minimum de modifications semble appropriée. Toutefois, l'algorithme de vérification et d'imputation à minimum de modifications de Fellegi-Holt devrait rester la méthode de choix pour les problèmes d'imputation de moins grande importance si l'on ne dispose pas d'un nombre suffisant de donneurs ou s'il est préférable d'utiliser une autre méthode pour effectuer l'imputation.



qui inverse l'ordre de l'opération, en commençant par rechercher des donneurs puis en modifiant le nombre minimal de variables, offre un avantage du point de vue des calculs. Toutefois, la NMI peut effectuer l'imputation en n'utilisant que des donneurs, tandis que la méthode de Fellegi-Holt peut être employée avec d'autres méthodes (telles que la méthode descendante). Statistique Canada a incorporé la NMI à son Système de contrôle et d'imputation du recensement pour les recensements de 2001 et de 2006.

10. Une série de réunions tenues après 2000 ont rassemblé de nombreux statisticiens européens qui se sont penchés sur divers aspects de la vérification et de l'imputation des données de recensements et d'enquêtes. Le débat a porté sur le thème de « La conception et l'évaluation de nouvelles méthodes de vérification et d'imputation » (EUREDIT) [Chambers, 2000; Poirier, 2000]. Statistique Canada a également conçu des méthodes de mesure et de réduction de la variabilité de l'imputation. Destiné aux enquêtes, le système SIMPVAR est conçu pour traiter les quatre principales méthodes d'imputation (ratio, moyenne, imputation dynamique et imputation par voisin le plus proche) [Rancourt *et al.*, 1997]. (Istituto Nazionale di Statistica) [ISTAT] a décrit le système de vérification et d'imputation mis au point pour le Recensement italien de 2001, utilisant le système DIESIS (Data Imputation and Editing System, logiciel italien) et d'autres méthodes (Di Zio, 2002; Bianci *et al.*, 2005).

11. En dehors de celles qui se fondent sur des cas réels, il existe d'autres méthodes d'imputation des données inconnues. On utilise parfois les mesures de moyennes. Certains pays utilisent des modèles de régression (Fédération de Russie, Comité d'État sur la statistique, 2000). La régression a été également utilisée pour imputer l'âge sur le formulaire abrégé du Recensement 2000 des États-Unis (Williams, 1998).

12. Grâce aux progrès de la saisie interactive, certains systèmes intègrent le codage, la saisie manuelle et la vérification en un seul système, en particulier pour les enquêtes. Tel est le cas du système CRIPTAX au Brésil (Hanono et Barbosa, n.d.) qui applique une méthode de vérification lors de l'introduction des données. D'autres systèmes, notamment le Système de traitement des recensements et des enquêtes [Census and Survey Processing System (CSPPro)], présentent certains aspects de vérification interactive. Comme on l'a indiqué précédemment, c'est aux bureaux nationaux de statistique de se prononcer sur la rentabilité des investissements qu'ils consacrent aux machines, aux logiciels, au personnel, en temps et ainsi de suite.

13. Comme pour la méthode descendante, le United States Census Bureau a conçu le Système intégré de traitement sur micro-ordinateur [Integrated Microcomputer Processing System (IMPS)] pour les recensements effectués de 1980 à 2000. Ce progiciel est sur DOS et comprend l'introduction, la vérification et la mise en tableaux des données, ainsi que d'autres fonctions, et de nombreux autres pays continuent de l'utiliser. À la fin des années 1990 et durant les années 2000, le Census Bureau a mis au point la version Windows de ce programme, appelé CSPPro (mentionné précédemment), qui effectue un grand nombre des mêmes tâches et est compatible avec Windows<sup>c</sup>. CSPPro fonctionne bien pour tous les recensements et enquêtes des petits pays; toutefois, le traitement est si lent que la vérification prend plus de temps qu'avec l'IMPS. Néanmoins, les pays qui commencent à effectuer leurs vérifications à partir de rien auraient intérêt à utiliser la version Windows. Les applications des systèmes IMPS et CSPPro peuvent toutes être conçues pour la méthode de Fellegi-Holt; la méthode par voisin le plus proche demanderait beaucoup trop de travail.

14. CSPPro permet à l'utilisateur de créer, de modifier et d'effectuer l'introduction des données, la vérification de séries et les tabulations dans un cadre de développement unique intégré. Ce programme traite les données au cas par cas (un ou plusieurs questionnaires), lorsqu'un cas consiste en un ou plusieurs enregistrements

c CSPPro est un logiciel d'introduction, de vérification, de tabulation et de diffusion de données de recensements et d'enquêtes qui combine les caractéristiques du Système intégré de traitement sur micro-ordinateur [Integrated Microcomputer Processing System (IMPS)] et du Système intégré d'analyse d'enquêtes [Integrated System for Survey Analysis (ISSA)] dans un environnement Windows.

de données. Les données sont stockées sur fichiers de texte ASCII (American Standard Code for Information Interchange) décrits par des dictionnaires de données. CSPro dispose d'un puissant langage commun pour l'application des règles de contrôle de l'introduction et de la vérification des données.

15. En particulier, la fonction de vérification des séries de données de CSPro permet d'identifier et de signaler la structure, la valeur et les défauts de cohérence des données des questionnaires. Le programme peut modifier (imputer) les valeurs des données sur la base de méthodes simples ou complexes. Il peut produire des rapports sommaires ou détaillés d'erreurs et de corrections, accéder à de multiples fichiers de consultation et lire des fichiers secondaires ou écrire dans de tels fichiers.

16. CSPro fournit également des outils pour visualiser des données et d'autres fichiers de texte, visualiser des tableaux et des cartes thématiques créés par CSPro, convertir des dictionnaires de données IMPS et ISSA en CSPro ou depuis CSPro, et convertir les fichiers (cartes) ESRI (Environmental Systems Research Institute) en fichiers de cartes CSPro. CSPro a été conçu conjointement par le U.S. Census Bureau, Macro International et Serpro, grâce à un important financement de l'USAID (United States Agency for International Development). CSPro est dans le domaine public, accessible sans frais et peut être distribué gratuitement; il peut également être téléchargé depuis le site [www.census.gov/ipc/www/cspro](http://www.census.gov/ipc/www/cspro).

## Glossaire

**Analyse rétrospective** (*Audit trail*) : Méthode permettant de suivre les modifications successives des valeurs attribuées à un champ de données, ainsi que la cause de chaque modification. Les analyses rétrospectives sont généralement effectuées une fois l'interview initiale terminée.

**Appariement (association)** [*matching*] : Selon la procédure d'imputation dynamique (*hot deck*), opération consistant à appairer un enregistrement donneur et un enregistrement destinataire (candidat) [Pierzchala, 1995].

**Appariement par fonction de distance** (*distance function matching*) : A pour effet d'attribuer la valeur d'une rubrique lorsqu'un enregistrement a été rejeté par une vérification, d'après un enregistrement dûment vérifié de l'unité voisine la plus proche, la proximité étant définie au moyen d'une fonction de distance, d'après les autres variables connues.

**Appariement statistique** [*statistical matching (in hot deck)*] : Opération qui consiste à appairer un enregistrement donneur et un enregistrement destinataire (candidat) selon un critère statistique déterminé, afin de pouvoir transférer les données du premier au second (Pierzchala, 1995).

**Cohérence interne** (*Internal consistency*) : Ce terme s'applique aux relations entre les variables d'une unité donnée de l'échantillon; c'est la cohésion interne que visent les vérifications dans la plupart des méthodes d'enquête (Ford, 1983; Pierzchala, 1995).

**Correction automatique** (*Automated correction*) : Corrections d'erreurs sur les données sans intervention humaine. Un des éléments de la vérification automatique des données (Pierzchala, 1995).

**Donnée fixe** (*cold deck*) : Matrice initiale d'imputation statistique. Base de référence des corrections effectuées, dont les éléments sont déterminés avant le début de la correction, et qui restent inchangés tout au long de son déroulement. Ainsi, la base de référence pourrait être constituée des données relatives à une année antérieure. Une méthode de « donnée » fixe modifiée peut consister à ajuster les valeurs de la donnée fixe en fonction des données du moment (données agrégées).

**Enregistrement** (*record*) : Représentation des données d'enquête stockées sur support magnétique et lisibles par ordinateur. Généralement, il existe un enregistrement par questionnaire, mais il est possible de scinder les données d'un questionnaire en plusieurs enregistrements, par exemple, de population et d'habitation (Pierzchala, 1995).

**Enregistrement « pur » (dont les erreurs ont été éliminées)** [*clean record*] : Enregistrement qui ne comporte aucune valeur manquante et qui a subi avec succès toutes les vérifications (Pierzchala, 1995).

**Enregistrements acceptés** (*passed edit records*) : Enregistrements qui ont été acceptés par toutes les vérifications concernant la rubrique en question.

**Enregistrements rejetés (par la vérification)** [*failed edit records*] : Lors des opérations de vérification et d'imputation, enregistrements rejetés par au moins une vérification concernant la (les) rubrique(s) en question.

**Erreurs de qualité** (*quality errors*) : Erreurs susceptibles d'affecter la qualité des données : par exemple, erreur systématique introduisant une distorsion (Granquist, 1984; Pierzchala, 1995).

**Fonction de distance** (*distance function*) : Dans le cas de données numériques fonction définie sur les variables homologues des enregistrements destinataires (candidats) et donneurs permettant de quantifier la notion de similarité. Cette fonction permet d'identifier des enregistrements appariés dans le cadre de la méthode d'imputation dynamique (*hot deck*) [Pierzchala, 1995].

**Imputation** (*imputation*) : Attribution d'une valeur à un champ de données, soit en cas de non-réponse, soit pour remplacer une valeur enregistrée dont l'incompatibilité avec une série de vérifications a été établie (Pierzchala, 1995).

**Imputation « hot deck » (à donnée variable)** [*hot deck imputation*] : Méthode d'imputation selon laquelle les enregistrements donneurs proviennent de la donnée momentanée de données de l'échantillon. [La méthode *cold deck* (à donnée fixe) désigne par contre la méthode d'imputation selon laquelle l'enregistrement donneur provient de données d'enquêtes antérieures] (Pierzchala, 1995).

**Imputation aléatoire à l'intérieur de classes** (*random imputation within classes*) : Usage de classes d'imputation afin de limiter la sélection aléatoire de l'enregistrement donneur à un ensemble censé présenter une certaine similarité par rapport à l'enregistrement devant faire l'objet d'une imputation.

**Imputation déductive** (*deductive imputation*) : Méthode selon laquelle une valeur manquante ou incohérente peut être établie avec certitude, souvent d'après les caractéristiques des réponses fournies à d'autres rubriques du questionnaire.

**Imputation déterministe** (*deterministic imputation*) : Cette situation est observée lorsque la valeur d'un seul champ permet à l'enregistrement de subir avec succès toutes les vérifications. Elle se produit dans certains cas (par exemple, lorsque la somme calculée des éléments d'un total n'est pas égale au total proprement dit). C'est la première solution à explorer en cas de vérification automatique et d'imputation des données d'enquêtes (Pierzchala, 1995).

**Imputation dynamique par donneur unique** (*single donor hot deck imputation*) : Méthode d'imputation des données à partir d'un donneur unique, pour un enregistrement rejeté, permettant d'imputer conjointement les valeurs de toutes les rubriques dans le cas d'un enregistrement identifié comme problématique par les vérifications.

**Imputation dynamique séquentielle** (*sequential hot deck imputation*) : Méthode d'imputation utilisée en cas de vérification séquentielle d'une série de variables, lorsque seules les valeurs vérifiées sont utilisées par la suite comme variables d'imputation dynamique.

**Imputation globale aléatoire** (*random overall imputation*) : Pour chaque enregistrement rejeté par la vérification, un enregistrement accepté est choisi au hasard dans l'ensemble d'enregistrements acceptés, et la valeur déclarée pour la rubrique en question est alors imputée à l'enregistrement rejeté par la vérification.

**Imputation multiple** (*multiple imputation*) : Cette méthode consiste à effectuer plusieurs imputations pour chaque valeur à imputer puis à donner une estimation de la valeur à affecter à la rubrique.

**Imputation par donneur** (*donor imputation*) : Méthode consistant à appairer chaque enregistrement destinataire ou candidat nécessitant une imputation avec un enregistrement issu d'une population « donneuse » définie comme, par exemple, dans le cas de l'imputation dynamique (*hot deck*) [Pierzchala, 1995].

**Imputation par la moyenne de classe** (*class mean imputation*) : Méthode utilisant des classes d'imputation définies pour créer des groupes d'enregistrements présentant une certaine similarité.

**Imputation par la moyenne générale** (*overall mean imputation*) : Attribue la moyenne de la rubrique tirée des enregistrements acceptés par les vérifications à la valeur manquante ou incohérente de tous les enregistrements rejetés.

**Imputation par voisin le plus proche** (*nearest neighbour imputation*) : Attribue à la valeur d'une rubrique d'un enregistrement rejeté celle tirée de l'enregistrement accepté du « voisin le plus proche »; la notion de proximité étant définie à l'aide d'une fonction de distance calculée d'après les autres variables connues.

**Imputation par modèle** (*model-based imputation*) : Utilise les données d'enregistrements acceptés par la vérification pour calculer par régression la variable devant faire l'objet d'une imputation sur la base d'un ensemble de variables prédictives.

**Imputation par régression** (*regression imputation*) : Utilise les données d'enregistrements acceptés par la vérification pour calculer par régression la variable nécessitant une imputation sur un ensemble de variables prédictives.

**Imputation statistique** (*statistical imputation*) : On pourrait choisir comme exemple d'imputation statistique l'utilisation d'un modèle de régression selon lequel la variable dépendante doit être attribuée par imputation aux variables indépendantes obtenues à partir de réponses présumées valides (Pierzchala, 1995).

**Imputation structurelle** (*structural imputation*) : L'imputation structurelle est utilisée lorsqu'une relation structurelle reste valable entre plusieurs variables. Par exemple, une valeur totale doit être égale à la somme des valeurs partielles (de ses éléments). Par conséquent, dans le cas d'une mère, le nombre d'enfants doit être égal à la somme des enfants vivants et des enfants décédés (Pierzchala, 1995).

**Indicateur** (*flag*) : Un indicateur est une variable servant à signaler des informations utiles concernant une ou plusieurs variables. Par exemple, en cas de modification d'une rubrique devenue valide après avoir été non valide, l'indicateur peut servir soit à enregistrer l'information initiale soit à signaler simplement la modification de la valeur de la rubrique.

**Introduction de données ininterrompue** (*heads down data entry*) : Mode d'introduction des données selon lequel la machine dans laquelle sont introduites les données ne détecte pas les erreurs au fur et à mesure de l'introduction de ces données, ce qui permet à l'opérateur de les introduire rapidement et sans difficulté.

**Jeu minimal** (*minimal set*) : Le plus petit ensemble de champs de données nécessitant une imputation propre à garantir que toutes les vérifications ont conclu à la validité des données (Pierzchala, 1995).

**Langage symbolique** (*pseudocode*) : Instructions ou spécifications de vérification écrites.

**Liste de codage** (*code list*) : Liste de toutes les valeurs admises (acceptables) susceptibles d'être attribuées à une rubrique.

**Macrovérification** (*macro-edit*) : Détection d'erreurs particulières à l'issue : 1) de contrôles effectués sur des données agrégées; ou 2) de contrôles effectués sur l'ensemble des enregistrements. Les vérifications sont fondées sur les estimations (Granquist, 1987; Pierzchala, 1995).

**Méthode de Fellegi-Holt** (*Fellegi-Holt method for automatic correction*) : Méthode de correction automatique ayant pour objet de modifier le plus petit nombre possible de rubriques de données; le modèle de Fellegi-Holt sert à identifier des jeux acceptables

de valeurs ou des gammes de valeurs pour les rubriques à imputer. On peut procéder par imputation séquentielle ou simultanée selon la méthode *cold deck* ou la méthode *hot deck*.

**Microvérification** (*micro-edit*) : Vérification traditionnelle effectuée sur des données au niveau de l'enregistrement. Antonyme logique des macrovérifications (Pierzchala, 1995).

**Micro/macrovérification** (*micro-macro edit*) : Procédure de vérification selon laquelle des microvérifications détaillées sont remplacées par une combinaison de microvérifications et de vérifications statistiques/macrovérifications. Les microvérifications effectuées selon la procédure mixte sont moins détaillées que dans le premier cas. Le principe consiste à « définir des vérifications d'enquête fondées sur la notion d'« incidence sur les estimations » et non sur une formule attrape-tout de détection des incohérences des données (Granquisit, dates diverses; Pierzchala, 1995).

**Nouvelle méthodologie d'imputation** (*new imputation methodology or Nearest-neighbour Imputation Methodology*) : Méthode semblable à la méthode de Fellegi-Holt dans la mesure où elle envisage simultanément toutes les vérifications, ne spécifie pas explicitement les opérations d'imputation et les effectue à partir d'un donneur unique. La méthode NMI examine chaque enregistrement rejeté de façon à déterminer les opérations d'imputation comportant le moins possible de modifications, sous réserve de la disponibilité de donneurs potentiels.

**Organigramme** (*flow chart*) : Description sous forme de diagramme de toutes les opérations qui doivent être exécutées.

**Passage machine ou phase d'exécution** (*production run*) : Opération consistant à traiter d'importantes quantités de données après avoir éliminé les erreurs initiales détectées par le programme de vérification ou d'exploitation.

**Pointeur** (*pointer*) : Variable utilisée pour marquer une rubrique ou une autre variable pour s'y référer ultérieurement. Par exemple, on utilise des pointeurs pour noter les numéros de ligne du « chef de ménage » et du « conjoint » pour utilisation ultérieure afin de s'assurer que les conjoints sont de sexe opposé et que l'un et l'autre sont mariés.

**Points aberrants** (*outliers*) : Valeurs de rubriques situées en dehors de certaines limites déterminées d'après certaines règles (Pierzchala, 1995).

**Pondération** (*weights*) : Selon la méthode de vérification et d'imputation proposée par Fellegi-Holt, des pondérations sont attribuées aux différents champs en fonction de la fiabilité des données. Plus le coefficient de pondération est élevé, plus l'imputation de la valeur de ce champs est probable (toutes choses étant égales par ailleurs). Des coefficients de pondération peuvent également être attribués aux vérifications (Pierzchala, 1995).

**Recherche** (*searching*) : Selon la procédure d'imputation dynamique (*hot deck*), opération consistant à rechercher un enregistrement donneur (Pierzchala, 1995).

**Saisie de données** (*data capture*) : Processus consistant à introduire les données recueillies sous une forme lisible par ordinateur. Des vérifications élémentaires sont souvent effectuées par des sous-modules du logiciel de saisie.

**Saisie discontinue de données** (*heads up data entry*) : Mode d'introduction des données selon lequel l'équipement d'introduction des données détecte les erreurs au fur et à mesure, ce qui permet à l'opérateur de les corriger immédiatement (Pierzchala, 1995). Voir « saisie interactive ».

**Saisie interactive** (*interactive keying*) : Mode d'introduction des données selon lequel l'équipement d'introduction détecte les erreurs dans les données au fur et à mesure, permettant à l'opérateur de les corriger immédiatement (Voir saisie discontinue de données).

**Sans réponse** (*non-response*) : Questionnaire incomplet ou manquant (Pierzchala, 1995).

**Série complète de vérifications** (*complete set of edits*) : Combinaison de vérifications explicites et implicites. Nécessaire au choix des domaines possibles d'imputation (si l'on veut que les valeurs imputées ne soient pas rejetées par les vérifications) [Pierzchala, 1995].

**Similarité** (*similarity*) : Dans le cas de données numériques, notion de proximité de deux enregistrements sur la base de variables d'appariement prescrites. Une fonction de distance permet de quantifier cette notion selon certains critères (Pierzchala, 1995).

**Système de Fellegi-Holt** (*Fellegi-Holt system*) : Se rapporte aux hypothèses et aux objectifs de vérification et d'imputation définis par Fellegi et Holt dans leur communication de 1976 au *Journal of the American Statistical Association*. Le modèle de Fellegi-Holt a pour trait essentiel de mettre en évidence la nécessité de recourir à des vérifications pour garantir que la présence d'une série de valeurs dans des champs de données qui ne sont pas imputées conduit toujours à l'obtention d'enregistrements définitifs (imputés) qui passent avec succès toutes les vérifications.

**Tableau ou rangée** (*array*) : Ensemble de valeurs numériques. Appelé parfois « matrice », un tableau peut servir à enregistrer des données numériques à caractère répétitif.

**Variables d'appariement** (*matching variables*) : Variables utilisées pour trouver une similitude entre un enregistrement destinataire (candidat) et un enregistrement donneur (Pierzchala, 1995).

**Vérification (1<sup>re</sup> définition)** [*edit (definition 1)*] : Limites logiques aux valeurs que peut prendre chaque variable (Pierzchala, 1995).

**Vérification (2<sup>e</sup> définition)** [*edit (definition 2)*] : Règles appliquées pour détecter des combinaisons de réponses inacceptables (Pierzchala, 1995).

**Vérification à plusieurs variables** (*multivariate edit*) : Type de vérification statistique utilisant des distributions à plusieurs variables pour évaluer les données et détecter les valeurs aberrantes (Pierzchala, 1995).

**Vérification de cohérence** (*consistency edit*) : Vérification des rapports de cohérence, par exemple de l'égalité entre un total et la somme de ses éléments, ou encore du fait que les « superficies moissonnées » sont toujours inférieures aux « superficies plantées » (Pierzchala, 1995).

**Vérification de validation** (*validation edit*) : Contrôle des vérifications effectué entre les champs d'un enregistrement particulier. Cela comprend le contrôle de chaque champ de chaque enregistrement pour s'assurer qu'il contient une information ou une donnée valide et le contrôle des informations contenues dans une combinaison prédéterminée de champs afin de s'assurer que les données enregistrées sont compatibles les unes avec les autres (Pierzchala, 1995).

**Vérification déterministe** (*deterministic edit*) : En cas de rejet, cette vérification met en évidence une erreur dont la probabilité est égale à un. Exemple : âge = 5 et état matrimonial = mère. Contraste avec la vérification stochastique (Pierzchala, 1995).

**Vérifications explicites** (*explicit edits*) : Vérifications explicitement établies par un spécialiste du sujet (les vérifications explicites sont définies par opposition aux vérifications implicites) [Pierzchala, 1995].

**Vérification implicite** (*implied edit*) : Vérification non formulée déduite logiquement des vérifications explicites établies par un spécialiste du sujet (Pierzchala, 1995).

**Vérification inter-enregistrements** (*between-record edit*) : Vérification effectuée sur des champs provenant de plusieurs enregistrements effectués dans le cadre d'une enquête. Les vérifications statistiques sont un exemple de vérification inter-enregistrements, car les distributions sont issues d'ensembles de champs de données tirés de tous les enregistrements effectués dans le cadre de l'enquête (Pierzchala, 1995).

**Vérification intra-enregistrement** (*within-record edit*) : Autre désignation des vérifications de validation (Pierzchala, 1995).

**Vérification manuelle** (*hand edit ou manual edit*) : Vérification effectuée par des agents avant introduction des données dans l'ordinateur (Pierzchala, 1995).

**Vérification rétrospective** (*edit trail*) : Voir « analyse rétrospective ».

**Vérification statistique** (*statistical edit*) : Série de contrôles fondés sur une analyse statistique des données fournies par les personnes interrogées : par exemple, le rapport entre les valeurs attribuées à deux champs se situe dans des limites déterminées par une analyse statistique de ce rapport pour des déclarants présumés valides (Greenberg et Surdi, 1984; Pierzchala, 1995).

**Vérification stochastique** (*stochastic edit*) : Dans un enregistrement rejeté, cette vérification met en évidence une erreur dont la probabilité est inférieure à 1 (Pierzchala, 1995).

**Vérification structurelle** (*structural edit*) : Contrôle fondé sur l'existence d'un lien logique entre plusieurs champs vérifiés. Par exemple, un total doit être égal à la somme de ses éléments; ou par suite d'un saut de champ inhérent à un questionnaire, deux variables situées sur deux « chemins » séparés ne peuvent être toutes deux non nulles. Une vérification structurelle garantit que la structure du questionnaire est maintenue dans l'enregistrement des données (Pierzchala, 1995).

**Vérifications quantitatives** (*quantitative edits*) : Vérifications appliquées à des champs de données mesurés sur une échelle continue (Pierzchala, 1995).



## Références

- Banister, J. (1980). Use and abuse of census editing and imputation. *Asian and Pacific Census Forum*, vol. 6, n° 3, p. 1 à 20.
- Bankier, M. (1999). Experience with the New Imputation Methodology used in the 1996 Canadian Census with Extensions for Future Censuses », Proceedings of the Workshop on Data Editing, UNECE, Rome (Italie).
- \_\_\_\_\_ (2005). Edit and imputation for the 2006 Canadian Census. Document présenté à la séance de travail sur la vérification des données statistiques, Ottawa, 16-18 mai 2005.
- \_\_\_\_\_, A.-M. Houle et M. Luc (n. d.). Canadian census demographic variables imputation. Manuscrit.
- Bankier, M., M. Lachance et P. Poirier (1999). A generic implementation of the new imputation methodology. Dans *Proceedings of the Section on Survey Research Methods*. Alexandria, Virginie, American Statistical Association, à paraître.
- \_\_\_\_\_ (2000). 2001 Canadian Census minimum change donor imputation methodology. Document présenté à la séance de travail ONU/CEE sur la vérification des données statistiques, Cardiff, Royaume-Uni, 18-20 octobre 2000.
- Bankier, M., P. Poirier, et M. Lachance (2001). Efficient methodology within the Canadian Census Edit and Imputation System (CANCEIS). Proceedings of the Annual Meeting of the American Statistical Association, 5-9 août 2001.
- Bankier, M., *et al.* (1996). Imputing numeric and qualitative census variables simultaneously. Dans *Proceedings of the Section on Survey Research Methods*. Alexandria, Virginie, American Statistical Association, p. 287 à 292.
- Bianchi, G. *et al.* (2005). New procedures for editing and imputation of demographic variables. Document présenté à la séance de travail sur la vérification des données statistiques, Ottawa, 16-18 mai 2005.
- Boucher, L. (1991). Micro-editing for the annual salary of manufacturers: what is the value added? Dans *Proceedings of the Annual Research Conference*. Washington, D.C., United States Bureau of the Census, p. 765 à 781.
- Chambers, Ray (2000). Evaluation criteria for editing and imputation in EUREDIT. Document présenté à la séance de travail ONU/CEE sur la vérification des données statistiques, Cardiff, Royaume-Uni, 18-20 octobre 2000.
- Chen, Bor-Chung (2007). CANCEIS Experiments of edit and imputation with 2006 Census test data. Statistical Research Division Study Series, n° 2007-1. Washington, D.C., U.S. Bureau of the Census.
- \_\_\_\_\_ *et al.* (2000). Using the DISCRETE Edit System for ACS Surveys. Bureau of the Statistical Research Division, Statistical Research Report Series, n° RR2000/03. Washington, D.C., U.S. Bureau of the Census.
- Commission de statistique des Nations Unies et Commission économique pour l'Europe (1994). *Statistical Data Editing*, vol. 1 : *Methods and Techniques*. Conférence

des statisticiens européens, Statistical Standards and Studies Series, n° 44. Numéro de vente : 94.II.E.36.

\_\_\_\_\_ (1997). *Statistical Data Editing*, vol. 2 : *Methods and Techniques*. Conférence des statisticiens européens, Statistical Standards and Studies Series, n° 48. Numéro de vente : 96.II.E.30.

De Waal, Tom, et Frank van de Pol (1997). A recipe for applying CHERRYPI in the edit process. Document présenté à la séance de travail sur la vérification des données statistiques, Prague, 14-17 octobre 1997.

Di Zio, M. (2002). Evaluating editing and imputation processes: the Italian experience. Document présenté à la séance de travail ONU/CEE sur la vérification des données statistiques, Helsinki, 27-29 mai 2002.

Fédération de Russie, State Committee on Statistics (2000). Data imputation based on regression models with variations of entropy. Document présenté à la séance de travail ONU/CEE sur la vérification des données statistiques, Cardiff, Royaume-Uni, 18-20 octobre 2000.

Fellegi, I. P., et D. Holt (1976). A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association*, vol. 71, n° 353 (Mars), p. 17-35.

Ford, Barry L. (1983). An overview of hot deck procedures. Dans *Incomplete Data in Sample Surveys*, vol. 2, *Theory and Bibliographies*. William G. Madow. Ingram Olkin et Donald B. Rubin, eds.

Granquist, L. (1984). Data editing and its impact on the further processing of statistical data. Document présenté au Workshop on Statistical Computing, Budapest, 12-17 novembre 1984.

\_\_\_\_\_ (1987). The short-term developing program for computer-supported editing at Statistics Sweden. Rapport présenté à la réunion du Data Editing Joint Group, Madrid, 22-24 avril 1987. Stockholm, Statistics Sweden.

\_\_\_\_\_ (1997). The new view on editing. *International Statistical Review*, vol. 65, n° 3, New York, Academic Press, p. 381-387.

\_\_\_\_\_, et J. G., Kovar (1997). Editing of survey data: how much is enough? Dans *Survey Measurement and Process Quality*, Lyberg *et al.*, directeurs de publication. New York, Wiley and Sons, p. 415 à 435.

Greenberg, Brian, et Rita Surdi (1984). A flexible and interactive edit and imputation system for ratio edits. Dans *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, Virginie, American Statistical Association p. 421 à 426.

Hanono, Reina Marta, et Dulce Maria Rocha Barbosa (n. d.). Generalized environment for application development for capturing, editing and coding statistical survey's data. Rio de Janeiro, Brazilian Institute of Geography and Statistics (IBGE).

Ireback, H. (2000). The impact of new information technologies on data collection at Statistics Sweden. Document présenté à la séance de travail ONU/CEE sur la vérification des données statistiques, Cardiff, Royaume-Uni, 18-20 octobre 2000.

Kalton, G., et D. Kasprzyk (1982). Imputing for missing survey responses. Dans *Proceedings of the Section on Survey Research Method*. American Statistical Association, p. 23 à 31.

- \_\_\_\_\_ (1986). The Treatment of missing survey data. *Survey Methodology*, vol. 12, p. 1 à 16.
- Kovar, J., et W. Winkler (2000). Comparison of GEIS and SPEER for editing economic data. Bureau of the Census Statistical Research Division Statistical Research Report Series, n° RR 2000/04. Washington, D.C., U.S. Bureau of the Census.
- Naus, J. I. (1975). *Data Quality Control and Editing*. New York, Marcel Dekker.
- Nordbotten, S. (1963). Automatic editing of individual statistical observations. Conférence des statisticiens européens, Statistical Standards and Studies, n° 2. New York, Organisation des Nations Unies.
- Organisation des Nations Unies (1992a). *Handbook of Population and Housing Censuses, Part I: Planning, Organization and Administration of Population and Housing Censuses*. Studies in Methods, Series F, n° 54. Numéro de vente : E.92.XVII.8.
- \_\_\_\_\_ (1992b). *Handbook of Population and Housing Censuses, Part II: Demographic and Social Characteristics*. Studies in Methods. Series F, n° 54. Numéro de vente : E.91.XVII.9.
- \_\_\_\_\_ (1999). *Standard Country or Area Codes for Statistical Use*. Statistical Papers, Series M, n° 49/Rev.4. Numéro de vente : E.98.XVII.9.
- \_\_\_\_\_ (2008). *Principles and Recommendations for Population and Housing Censuses, Revision 2*. Statistical Papers, Series M, n° 67/Rev.2. Numéro de vente : E.07.XVII.8.
- Pierzchala, M. (1995). Editing systems and software. Dans *Business Survey Methods*. B.G. Cox, *et al.*, directeurs de publication. New York, John Wiley and Sons.
- Poirier, C. (2000). A prototype knowledge base on data editing and imputation. Document présenté à la séance de travail ONU/CEE sur la vérification des données statistiques, Cardiff, Royaume-Uni, 18-20 octobre 2000.
- Pullum, T. W., T. Harpham et N. Ozsever (1986). The machine editing of large-sample surveys: the experience of the World Fertility Survey. *International Statistical Review*, vol. 54, p. 311 à 326.
- Rancourt, E., *et al.* (1997). Estimation of variance in presence of imputation. Proceedings of Symposium 1997: New Directions in Surveys and Censuses. Ottawa, Statistique Canada, p. 273 à 279.
- Rubin, D. B. (1987). *Multiple Imputation for Non-response in Surveys*. New York, Wiley.
- Sande, I. G. (1982). Imputation in surveys: coping with reality. *American Statistician*, vol. 36, p. 145 à 152.
- Särndal, C. E., B. Swensson et J. Wretman (1992). *Model Assisted Survey Sampling*. New York, Springer-Verlag.
- Statistique Canada (1998). *Statistics Canada Quality Guidelines*, 3<sup>e</sup> édition. Ottawa, Statistique Canada.
- Williams, Todd R. (1998) Imputing person age for the 2000 Census Short Form: a model-based approach. Washington, D.C., Bureau of the Census Statistical Research Division, Statistical Research Report Series, n° RR98/07.

- Winkler, W. E. (1997a). Edit/imputation System for the U.S. Decennial Census. Document présenté à la séance de travail sur la vérification des données statistiques, Prague, 14-17 octobre 1997.
- \_\_\_\_\_ (1997b). Set-covering and editing discrete data. Technical report. Washington, D.C., U.S. Bureau of the Census.
- \_\_\_\_\_ (2006). Data quality: automated edit/imputation and record linkage. U.S. Census Bureau Statistical Research Division Research Report Series, n° 2006-7. Washington, D.C., U.S. Bureau of the Census.
- \_\_\_\_\_, et B. C. Chen (2002). Extending the Fellegi-Holt Model of Statistical Data Editing. U.S. Bureau of the Census Statistical Research Division, Statistical Research Report Series, n° 2002-02.
- Winkler, W. E., et L. R. Draper (1997). The SPEER edit system. Dans Commission de statistique des Nations Unies et Commission économique pour l'Europe, *Statistical Data Editing*, vol. 2, Methods and Techniques. Statistical Standards and Studies, n° 48. Numéro de vente : E.96.II.E.30, p. 56 à 62.