



**Economic and Social
Council**

Distr.
GENERAL

CES/2001/28
17 January 2001

ORIGINAL: ENGLISH

**STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

Forty-ninth plenary session
(Geneva, 11-13 June 2001)

Report of the October 2000
Joint ECE/WHO preparatory Meeting on Measuring Health Status

I. INTRODUCTION

1. Following the recommendation of the 1998 Joint ECE/WHO Meeting on Health Statistics in Rome, the Conference of European Statisticians decided to convene a preparatory meeting together with WHO and a small number of interested countries to try to make progress towards a common framework for measuring population health. Statistics Canada kindly offered to host the meeting which took place in Ottawa 23-25 October, 2000.

2. The meeting was opened by Mr. Ivan Felligi, Chief Statistician, Statistics Canada. He told the participants that the challenge to countries is to be willing to collaborate in the development of international standards, to be willing to adopt them when they exist (even where this risk disrupting well established time series), and last but not least, to be ready to benefit from all this effort through learning experiences of other countries.

II. ORGANISATION AND DOCUMENTATION

3. The meeting was attended by 41 experts from the following countries Australia, Canada, Denmark, France, Italy, Netherlands, France, New Zealand, Spain, United Kingdom and United States. Eurostat was also represented. Representatives from the OECD, World Bank, Inter American Development Bank, WHO-EURO, WHO-PAHO and the United Nations Statistics Division also

participated.

4. The provisional agenda was adopted and the nine sessions were chaired by the following persons: Michael Wolfson (Statistics Canada), Alan Lopez (WHO), Gary Catlin (Statistics Canada), Ed Sondik (National Center for Health Statistics, United States), Richard Madden, (Australian Institute of Health and Welfare, Australia), Marijke de Kleijn-de-Vrankrijker (Prevention and Health, Netherlands), Gouke Bonsel (University of Amsterdam, Netherlands).

5. The agenda was discussed on the basis of 26 submitted papers and covered the following topics:

- (i) Uses and needs for population health status measures;
- (ii) National initiatives;
- (iii) International agencies' initiatives;
- (iv) Framework for measuring population health;
- (v) Critical review of the comparability of surveys and data;
- (vi) National survey data to assess domains for standardized health status measures;
- (vii) Comparability of health status assessment methods;
- (viii) New methodological approaches towards cross comparability; and
- (ix) Perspectives on health status assessment.

6. All papers, including the report, are available at the following Website address:

www.unece.org/stats/documents/2000.10.health.htm

III. MAIN CONCLUSIONS AND RECOMMENDATIONS

7. The meeting concluded that enhancing cross-population comparability of health status is important. Indeed it raises important issues for the use of health status data within countries, especially those with ethnically or culturally diverse populations. Current approaches to cross-population comparability need strengthening, and warrant further research and development.

8. Health status measurement is insufficient by itself; it needs to be embedded in a framework that allows health status assessment to be meaningfully connected to health policy. The WHO framework is a major step towards providing such links. A survey instrument alone is not sufficient: recommendations concerning additional operational issues, for example sampling approaches and administrative processes should also be further developed. Furthermore, the process of compiling and aggregating data, for example into summary measures of population health or other formats, should be transparent. This requires that the underlying data and clear descriptions of methods are widely available.

9. Based on a small survey of the attendees, a majority of those responding (n = 23) appeared comfortable with the candidate list of domains suggested by WHO to be used within its framework as a basis for a standardized health status description, subject to two main adjustments: (a) lower priority

should be given to digestion, bodily excretion, skin & disfigurement and sexual activity; and (b) more detailed domains on psychological health should be developed (as also supported by the Canadian analysis presented, see discussion under Session 6). At the same time, the rest of the respondents (n = 10) did not respond to this part of the questionnaire on WHO's proposed domains, though they did submit a range of other comments.

10. Several references were made to special populations, especially children and frail elderly, and the need for the health status measures to be appropriate for these groups.

Recommendations

11. WHO was requested, both in general and in relation to the specific draft resolution for the World Health Assembly, to intensify its consultation with member states and appropriate expert groups, and to build further on the work done both within countries and within international organizations and research groups.

12. Building on the substantial progress already made by WHO and the need for ensuring a continuous country participation in the further development of the Common Framework, the meeting recommended that Statistics Canada investigate the possibility of constituting a small group of interested WHO Member States to collaborate with WHO on those aspects of the framework requiring further research.

13. The meeting asked the UN/ECE to submit its conclusion to all the members of the Conference of European Statisticians and asked the CES to convene the planned formal inter-governmental meeting in 2001/2002 together with WHO, to consider, among other agenda topics, the adoption and application of the WHO Common Framework on Measuring and Reporting on the Health of Populations, a version of which is expected to be adopted by the WHO in 2001.

14. The above conclusions and recommendations were agreed upon in the last session and adopted by the participants.

ANNEX

SUMMARY OF DISCUSSION

Opening Opening note from Lene Mikkelsen (UN/ECE) and Chris Murray (WHO). Official opening statement by Dr. Ivan Fellegi, Chief Statistician.

Session 1: The uses of and needs for population health status measures

Chair – Michael Wolfson, Statistics Canada

Speakers:- Ed Sondik, National Centre for Health Statistics, U.S.; Michael Decter, Chairman of the Canadian Institute for Health Information; John Fox, Chief Statistician, Department of Health, UK; Marleen de Smedt, Eurostat

1. The speakers addressed the uses of and needs for health status measures. Key areas stressed across presentations included the need for international comparability and their direct input to decision-making. They also stressed that much of the demand for health information is local, but that this must be comparable to information at regional, national and cross-national levels. For example, local area measures are essential for policy and should be estimated through appropriate sampling or modeling techniques.
2. The demands for information on health status, as well as the determinants of health status, highlight that cross-sectional data are not sufficient -- longitudinal measures that provide information on transition probabilities and the dynamics of health and its correlates, are also needed.
3. Health status measures should assist in the evaluation of the effectiveness of interventions, thereby enabling more effective application of policy. One challenging area is inequalities in health. It was stressed that further research has to be undertaken to improve methodologies that assess changes in health status.
4. In addition, better measures that communicate investment in health, other than the most widely used health statistic, the “percentage of GDP spent on health care”, are needed for enriched policy discussions. Overall summary measures of population health that enable cross-national comparisons should be constructed so as to be readily disaggregated by age, cause and population subgroups. Finally, the development of cross-nationally comparable data should build on international harmonization efforts.

Session 2: Roundtable on National Initiatives

Chair: Alan Lopez, WHO

Speakers: Richard Madden, Australian Institute of Health and Welfare (AIHW); Gary Catlin, Statistics Canada; Vittoria Buratti, Italian National Statistical Institute (ISTAT); Niels K. Rasmussen, Danish Institute for Clinical Epidemiology (DICE); Siobhan Carey, Office of National Statistics UK (ONS);

Martin Tobias, Ministry of Health of New Zealand; Jean-Louis Lanoe, Institut national de la statistique et des études économiques (INSEE) ; Jennifer Madans / NCHS, U.S.; Marthe Gold / NYU, U.S.

5. The speakers presented a series of concise overviews of recent developments and initiatives regarding measuring population health, including household survey program developments and their links to health goals and / or population health monitoring. This was followed by a plenary discussion where the following issues were raised:

- **Need comparable operational definitions.** Definitions of ill health or disability vary considerably across countries and within countries over time. Such differences may lead to variations in the threshold definitions of individuals considered disabled or at some minimum level of ill health. For instance, Australia employs a large number of domains to define disability in surveys, and this may broaden the number of people assessed as disabled, in comparison to other countries which utilize narrower inputs. Although there was no clear answer to this issue, it is clear that in order to compare overall levels and trends in disability, it is necessary to ensure that definitions are comparable. Unfortunately, international comparisons in the past have rarely been a priority for countries, although all countries are interested in consistent time series data, and some countries are interested in comparisons with neighboring countries. In general, including health status data, beyond the self-report of morbidity, was a welcomed trend within countries, and was specifically discussed within the evolution of French surveys.
- **Facilitate sub-group comparisons.** Most countries would like to be able to compare sub populations but are hampered due to small sample sizes that do not allow for sub-population estimates. Different countries provided examples where specialized sub-populations or geographic areas were over-sampled in order to obtain estimates. For example, in Canada, a new Community Health Survey with a large sample of 130,000 has been introduced specifically to generate data for smaller sub-provincial regions. It will also be possible to compare socio-demographic and ethnic groups in considerable detail.
- **Broaden sampling frame.** Most surveys are representative of the non-institutionalized civilian population. Few countries conduct nationally representative surveys of the institutionalized population. The question raised was if the population in institutions should be included, particularly to increase the comparability of survey data across countries. Those countries which had done so found it very worth while, particularly for some sub-populations, i.e., 75 years of age or older. In a related point, non-response by selected groups was discussed. Country experience varied greatly. For example, in the US, it is estimated that individuals with disabilities were well covered within the National Health Interview Survey, probably due to the fact that responding and being counted was in their self-interest. The democratic nature of surveys was highlighted: an inclusive sampling frame and well-representative survey should not only be viewed as a measurement tool but as a way for individuals to communicate.
- **Facilitate local area estimates.** There was general agreement that in addition to sub-group comparisons, approaches to obtain local area estimates should be further developed and tested. For example, in some countries local level estimates are obtained through sampling (i.e. Canada) and in others (i.e. United States) through modelling. The validity of different approaches, and comparative costs associated with obtaining these estimates should be evaluated.
- **Develop criteria for good measurement tools.** Several countries described selected standardized health status instruments as valid or appropriate, although the criteria to make such evaluations were not made explicit. For example, various opinions were expressed on the appropriateness of the SF-12 or SF-36 in different populations and sub-populations, with some citing convergent

validity, and others citing differences in the underlying principle components structure (psychometric analysis) within different sub-populations. It was suggested that one criteria for adopting or switching among standardized instruments should be improvement in the data's international comparability. Further work on establishing criteria for validity and reliability within a population, and on cross-population validity are required, that reflect scientific advances in the measurement of health status. Regulatory mechanisms, such as legislation in New Zealand which requires that health surveys take place regularly, was also cited as a good practice in the institutionalization of measuring health status in the population.

- **Increase access to unit record data.** The need to facilitate access to micro data for secondary analysis by the research community is generally recognized. However, possibilities differ according to countries' data protection laws and traditions. For example, rich micro-data sources in Denmark and other Scandinavian countries can be accessed by researchers through a collaborative basis. In the UK there is a longitudinal link to census data, as well as a centre that facilitates access to data collected by the national statistical agency, regions and other clinical data base. Several cross-national collaborative efforts, such as those that have tested standardized instruments in a wider range of countries, should make available the data collected.

Session 3 : International Agencies' Initiatives

Chair: Gary Catlin, Statistics Canada

Speakers:, Marleen De Smedt, Eurostat; Pieter Kramers, National Institute of Public Health and the Environment; Anatoli Nossikov, WHO/EURO; Norberto Dachs, WHO/PAHO; Manfred Huber, OECD; Margaret Mbogoni,UNSD; Margaret Rothman, Johnson and Johnson

6. Each presenter gave a brief overview of the relevant health measurement work their organisation/companies were undertaking. As this was supposed to be an information session there was no general discussion afterwards. Nevertheless, some of the key points raised in this session included:

- 1) the importance of consensus building in the development of comparable surveys cross-nationally;
- 2) the need to provide methodological tools and guidance to government statistical institutions and regions, including model surveys; and
- 3) the recognition that self-reported health is increasingly being used as an additional way to evaluate technologies and interventions within health services, covering both public and private sectors.

7. The common theme across presentations was the need for further conceptual agreement and methodological work that developed standardized tools that took into account cross-population or cultural differences.

Session 4: Toward a Framework for Measuring Population Health Status

Chair: Ed Sondik, United States

Speakers: Michael Wolfson, Statistics Canada; Chris Murray, Bedirhan Üstün, Alan Lopez, all from WHO Global Program on Evidence for Health Policy, Geneva

8. A brief introduction noted by way of background the joint ECE/WHO meeting on health statistics in Rome in October 1998 as the catalyst for this meeting, and that the WHO initiatives being presented in this session could provide an opportunity to move ahead more expeditiously on the task of developing internationally comparable measures of health status.

9. The WHO common framework for measuring and reporting information on the health of populations was presented. It includes three components that build on major advances in measuring health status, and address key challenges facing this field. These three components include:

- 1) The WHO Family of International Classifications on Health (e.g. including ICD and ICIDH and revision processes).
- 2) Operational systems of data collection on Health States of Populations (e.g., standardized health status instrument development and testing in nationally representative cross-national surveys)
- 3) Summary Measures of Population Health (e.g., health expectancy measures, such as healthy life expectancy and health gap measures, such as disability-adjusted life expectancy).

10. The presentation provided the historical context of this framework, followed by a review of the process to develop each of the three components, with an emphasis on the second component, relevant to the objectives of this conference. In order to develop and discuss the framework, in particular the second component, WHO held extensive consultations over the past six months leading to this conference, including the hosting of two working group meetings in Geneva made up of several individuals from different Member States, many who were participants at this conference.

11. The current survey instrument under development builds upon existing instruments as well as additional qualitative work in a large range of countries. The survey is currently being tested in more than 10 nationally representative surveys. Data collected will be compatible with the ICIDH and contribute to the calculation of summary measures of population health.

12. Furthermore, the basis for the recommendations concerning the WHO common framework, made by the WHO Committee of Experts on Measurement and Classification for Health in September 2000 and currently under consultation within all of WHO's 191 Member States, was also presented for discussion. An eventual resolution in this area that is endorsed by the WHO Executive Board and by the World Health Assembly will enhance the institution's mandate to provide comparable data on the health of populations cross-nationally, in conjunction with Member States.

13. This presentation was followed by a plenary discussion where participants asked for clarification as to the process and future plans of WHO in the area of health status measures. While the active leadership role of WHO was welcome, it was felt that the consultation with experts in member countries should be further intensified. While it was not a focus of the session, ICIDH-2 also generated important discussion.

14. The following issues were discussed:

- **Health status measurement – one of three legs of the common framework.** The second component of the WHO three-part framework consists of a measurement strategy including the development and testing of a standardized instrument / series of questions for use on health interview surveys. The health profile under consideration covers 21 health domains selected

from the most recent version of the ICIDH. Information collected through surveys incorporating this instrument may also be used to construct summary measures of population health. By integrating the methods and objectives of all three components of the framework, one of the basic questions that should be answered is whether population A is healthier than population B, and what are the contributing causes.

- **Comparable information on health across populations.** Several issues and objectives regarding comparable population health status information were raised in the discussion. It was observed that the development of common standards for health measurement is a truly daunting but necessary task. Currently, we are able to compare what we spend on health (i.e., as a percentage of GDP) but not how healthy we are – a profound gap in health information.
- **Standardized health status description, i.e., comprehensive set of domains to describe health.** A multi-dimensional health status profile should include enough domains to describe a broad array of health states. Participants agreed that a clear, structural model for health is needed. During the meeting, the domains used for individual level description as a basis for the operational measurement of health were discussed in detail. It was also noted that as part of the third component of the WHO framework, summary measures of population health (SMPHs), six of the 21 domains were singled out to form a shorter profile as the basis of SMPHs.

15. In support of this discussion, Statistics Canada prepared a questionnaire with a listing of the 21 domains and their definitions, and participants were asked to comment on their appropriateness. In addition to the two dozen or so domains for health status profiling, the questionnaire also asked participants which eight domains as priority domains for the shorter profile to be used as the basis for SMPHs. An analysis of the results of this survey were circulated and discussed the next morning.

16. In sum, the results suggest that the top 10 domains among the 21 were (in roughly descending order): mobility, self-care, usual activities, cognition, pain, affect (with suggestions for disaggregation), social functioning, communication, and vision. For the subset that should form the core for summary measures of population health, the leading domains were (in no particular order): pain, affect, cognition, mobility, self-care and social functioning; while the next group (again in no particular order) were: vision, hearing, energy/vitality, and communication.

17. The question was also asked how WHO had arrived to the current list and what had been sacrificed to get the list? The meeting was told that the current domains had been discussed and agreed upon by other expert groups earlier this year in Geneva, and were originally based on a review of some 300 assessment instruments. Groupings of items had been made by several expert meetings and qualitative feed back from collaborating centers around the world.

18. Regarding the number of priority domains for SMPH use, it was pointed out that with approximately four domains most of the variance in relation to health state valuation would be obtained; additional domains may add little significant information. Yet, it is important to develop a comprehensive list of domains that allow for a broad range of health states to be described adequately. Leaving out domains implies that the importance of that domain is zero. Whether any domains are missing was also discussed.

19. For SMPH and many practical purposes, 21 domains (the candidate list proposed by WHO) are too many – in fact none of the most widely used health status profile measures currently developed for description or valuation purposes have 21 domains, i.e. EUROQOL, SF-36 or HUI-III. Therefore reducing the number of domains, based on conceptual discussions and empirical investigations, should be pursued. The final choice of domains should be based on evidence that: there is comparability across

countries, that measures reach acceptable levels of reliability and validity, and they may be interpreted in conjunction with additional information such as clinical measures or socio-demographic characteristics.

20. The following issues were discussed at some length:

- **Update standardized health state descriptions and develop external tests.** It was agreed that the list of domains which was chosen would not represent a definition of health but rather what could be used as a basis to compare health cross-nationally for some period in the near future, i.e., next 5 years, and then should be reviewed. Some domains directly measure health status, i.e., vision, while others serve as an indirect description of health status, i.e., usual activities. Developing valid and reliable external tests for each domain present different challenges. An external validation of questionnaire responses for usual activities is much harder to develop than a test of good vision, in particular one that is cross-culturally valid.
- **Connection with other international initiatives.** Some questions were raised about the connection with the other ongoing initiatives particularly within the European region – some of which seem to overlap. Regarding European initiatives, these differ from the WHO framework and some have different objectives. For example, in the EURO HIS initiative, the 12 categories or domains covered include survey modules on risk factors, which are distinct from health status, and there is limited overlap on those domains describing health status: only self-reported health, and long and short-term disability overlap. Within self-reported health, only general health is covered and there is no further breakdown of domains. More importantly, however, the WHO framework goes beyond reviewing differences in the wording of existing surveys, or suggesting survey questions for specific domains. It also proposes a common conceptual basis to describe health, and includes in its assessments of cross-population validity not only translation between languages, but also whether the meanings associated with responses in different cultural contexts are comparable. The WHO framework includes an operational approach that covers a broad range of domains that describe health, and is implementing nationally representative surveys in a large number of countries to further refine methods to improve cross-population comparability.
- **Process to implement framework.** While the structure, development of instruments and summary measures are all-important and need to be developed across countries, discussion also suggested that more emphasis should be put on the process and long term goals and less on short-term goals.
- **Summary measures.** Summary measures of population health were also briefly discussed. It was agreed that these were excellent to get an overview of the situation, and that the underlying data provides further insights on the causes of the disease burden. It was also pointed out that while summary measures are essential to the reporting of health they are not intended as measurement tools.
- **Policy use.** The question was raised what the policy utilization of this instrument would be; for example, how is the data collected through a standardized instrument going to lead to a better understanding of cardiovascular disease. WHO replied that the framework is to be seen as one step in a comprehensive process. The first step is to be comparable on describing health status. Other aspects, such as modules assessing the determinants or risk factors of health are being developed by other programs within WHO that focus on specific risk factors or disease groups,

which may be considered as complementary to the common framework initiative. Furthermore, attention was drawn to the reality that how study results are publicized and disseminated influences to a large degree whether results raise the interest of decision makers at various levels. Providing appropriate information to the media has been under utilized and should be pursued.

Session 5 : Critical review of the Comparability of Surveys and Data

Chair: Richard Madden, AIHW

Speakers: Marleen De Smedt, Eurostat; Gaetan La Fortune, OECD; Niels Rasmussen, representing EURO-REVES; Ritu Sadana, WHO

21. Each presenter provided a distinct perspective on the comparability of surveys on health status. The first two presentations focused on differences in the wording of questions across surveys, within the EU and OECD countries respectively, and concluded that further efforts are required to improve the comparability of surveys. The third presentation focused on this point by highlighting that “comparability of items is a key criterion for questionnaire development,” with specific reference to the development of model questionnaires for use within Denmark. A dominant theme across the first three presentations was that international cooperation to improve comparability must be further strengthened, even if perfect comparability is unobtainable.

22. The fourth presentation introduced an additional challenge for the meaningful comparison of household interview data across populations. Beyond the use of different questions and response scales to assess health, limitations in the cross-population comparability of data exists as surveys with only self-reported data lack external criteria to calibrate responses across countries. The analysis presented represents an initial effort to address these challenges by developing and testing a simple analytical method to compare responses to different questions across surveys based on an analysis of unit level data and using only self-reported responses. Despite attempts to improve comparability, the valid comparison of data across countries appears limited due to non-random reporting differences even where the survey methodologies and data collection approaches are well standardized.

23. Given that surveys can be relatively inexpensive and are widely available from a growing number of countries, the presentation and subsequent plenary discussion concluded that further research to develop and test methods to improve the cross-population comparability of survey data, based on external calibration approaches, would be worthwhile. The discussion noted that “bias” is an inappropriate term for the lack of cross-population comparability; non-random reporting differences is offered as an alternative, less judgmental term.

Session 6 : Using National Survey data to Assess Domains for Standardized Health Status measures

Chair: John Millar, CIHI

Speakers: Siobhan Carey, ONS, UK; Jean-Marie Bethelot and Julie Bernier, Statistics Canada

24. The first presentation provided a review of six survey instruments implemented within England over several years, if not decades, based on established criteria, empirical comparisons based on

nationally representative survey data, and cognitive testing of selected questions. Differences and/or similarities in the prevalence of different health states (single item questions or domain scores), based on the range of instruments compared, were partially explained by differences in the interpretation of questions, the different cognitive processes people used to respond to questions, whether individuals reported having a long-standing illness or not, and the relatively recent use of health services. Many of these differences contribute to the lack of within population comparability of data.

25. The second presentation focused specifically on the candidate list of 21 domains proposed within the WHO common framework. Using both qualitative reviews and quantitative analytical methods, the analysis assessed whether existing standardized instruments or modules included in nationally representative household surveys in Canada already collect information on these domains. The analysis also explored whether additional domains should be added to the WHO list in order to reflect better the main kinds of variation in health status observed in the Canadian surveys. Seven major composite domains of health from the Canadian surveys were identified using exploratory common factor analysis: psychological well-being, stress, functional limitations, chronic conditions, disability days, sensory impairment and depression. Additional domains that contributed unique information, over and above the composite factors, included breathing, hearing, energy expenditure, emotion, thinking, social support, and incontinence.

26. The analysis observed that most domains proposed by the WHO candidate list are covered within the Canadian surveys. The WHO domains generally seem to capture lay persons' definitions of health, with the important exception of "health reserve". However, "affect" within the WHO candidate list appears too general and vague. It comprised three distinct domains in the Canadian data (psychological well-being, stress, depression) and these should be included within the WHO candidate list. The analysis concluded that further work on developing additional domain(s) should be pursued.

Session 7 : Comparability of health status assessment methods

Chair: Marijke de Kleijn-de-Vrankrijker

Speakers: John Ware, Quality Metric, Inc., U.S.; Alan Tenant, University of Leeds, UK; Gouke Bonsel, University of Amsterdam, Netherlands

27. Given the variety of challenges to compare data on health across populations, the main theme of this session was to review the current state of the art on classical and modern psychometric approaches to data analysis and comparative interpretation within the health status measurement field.

28. The main presentation in this session made the following major points:

- 1) classical and modern psychometric methods complement each other;
- 2) scales measuring the same concepts and translations of instruments can be calibrated on a common health metric sufficiently to compare results, using item response theory analysis;
- 3) items with known properties across populations can form the basis of comparable cross-national instruments; and

- 4) a new conceptual model of health confirmed by empirical testing is necessary for international research.

29. The two discussants provided further insights on modern methods, i.e. Rasch and IRT (item response theory) analysis. One use is to identify differential item functioning (when survey questions do not provide comparable results across or within populations). The discussants highlighted the (strong) assumptions necessary to conduct such analysis (e.g. that the underlying concept is uni-dimensional), and other limitations of IRT analysis. In particular, point 2 above was questioned as to whether calibration techniques relying only on internal approaches to establish a common metric across populations were in fact reliable and valid. The discussion underscored that although these new methods contribute to improved comparability, many analytical variations exist, and their appropriate use is widely discussed and debated.

Session 8 : New methodological approaches towards cross comparability: external calibration

Chair: Gouke Bonsel

Speaker: Chris Murray, WHO Global Program on Evidence for Health Policy, Geneva

30. The presentation introduced that cross-population comparability refers to the property that a similar response on an item or domain score maps to the same true level of health in different populations. In addition to the classical psychometric properties of instruments (e.g. reliability and validity) and other issues concerning survey implementation (e.g., sampling frame, missing data), the issue of cross-population validity emerges as an essential requirement for international comparisons, beyond whether the same instrument and questions are used among different populations. Even though some of the assumptions required to use IRT models are often violated when applied to the measurement of health status, as discussed in the previous session, efforts to build on this approach to evaluate the comparability of responses appear worthwhile. Nevertheless, without external criteria, it was argued that current efforts provide insufficient evidence to establish the cross-population comparability of data.

31. Towards building on current efforts and enhancing methodological work in this area, WHO plans a comprehensive strategy to enhance the comparability of data from household interview surveys. Two specific areas of work are currently being pursued through international consensus and research with Member States and collaborators. These are:

- 1) identifying and testing a core set of domains to describe health and appropriate survey questions for household interviews, and
- 2) developing and testing methods to estimate cross-population comparability of data collected and to calibrate self-reported response categories across populations to enhance comparability.

32. Based on initial results from pilot surveys, one method to calibrate responses across populations, namely comparing self-reported health and tested health (performance tests) in selected domains, such as vision, was discussed in greater detail. Preliminary results and challenges were commented on in the plenary discussion in the next session.

Session 9 : Perspectives on Health Status Assessment

Chair: Michael Wolfson, Statistics Canada

Panel Members: Ed Sondik, USA and John Fox, UK

33. The introductory comments to this session highlighted the following points:

- Comparability of data is an issue both across and within populations;
- Health status measures and survey results must first be useful within countries, and then as a secondary priority provide internationally comparable data;
- The new methodological procedures introduced by WHO are promising and are worthy of further development and testing;
- Additional challenges to health status assessment include the measurement of change, inequalities, and child health measures, among a broader consideration of other factors that determine health or are related to health;
- The review of various activities and approaches demands further coordination among international agencies.

34. The subsequent discussion focused on the need for further guidance on methodological advances and the systematic collection of evidence on how to test and adjust for cross-population comparability. Making existing data and future nationally representative data available to this research effort is a necessary step, despite current barriers to accessing data for secondary analysis.

35. The discussion also recognised that the general approach described by WHO in its common framework is worthwhile pursuing and that the linked research agenda around the empirical basis for measuring and comparing health status deserves to be implemented. With the aim to further that and to promote closer cooperation among the international agencies, the European Union represented by EUROSTAT offered to host a workshop on current data collection and harmonization efforts in light of the WHO framework.