UNITED NATIONS **E**

Economic and Social Council

Distr.
GENERAL

ECE/CES/2006/4/Add.2
1 February 2006

Original: ENGLISH

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

Fifty fourth plenary session
Paris,12-16 June 2006
Item 4 of the provisional agenda

Report of the Meeting of the Group of Experts on Measurement of Health Status, held on 14-16 November 2005

Note by the secretariat

This meeting was organised jointly with World Health Organization and Eurostat.

INTRODUCTION

1. The Joint UNECE/WHO/Eurostat Meeting on Measurement of Health Status was held on 14-16 November 2005 in Budapest, Hungary. It was attended by participants from Australia, Austria, Bulgaria, Canada, Czech Republic, Denmark, Finland, France, Hungary, Ireland, Italy, Latvia, Lithuania, Mongolia, Netherlands, New Zealand, Norway, Poland, Republic of Moldova, Republic Of San Marino, Romania, Russian Federation, Serbia and Montenegro, Slovak Republic, Slovenia, Spain, Switzerland, Turkey, and the United States of America. Country representatives were from National Statistical Offices, Ministries of Health, Public Health Institutes, and other national institutes concerned with health and statistics. The European Commission was represented by Eurostat and assisted by colleagues from the EU Partnership Health. The World Health Organization (WHO), the International Labour Office (ILO), and the Organization for Economic Cooperation and Development (OECD) were also present. Three experts, one from University of Ottawa (Canada), one from the University of Amsterdam (the Netherlands), and one from the PROMIS Network (USA), participated at the invitation of the

secretariat.

2. Ms. Jennifer Madans (USA) was elected as Chairperson of the meeting.

3. The following substantive topics were discussed during the meeting sessions based on the 11 invited papers:

Session 1: Background, Terminology, and Scope
Presentations given by WHO, Australia, and the USA.
Discussants: Mr. Michael Wolfson (Canada), Mr. Howard Meltzer (EU Partnership Health/UK), and Mr. Arpo Aromaa (Finland).

Session 2: Criteria for Selecting Domains and Identification of Domains
Presentations given by Canada and the PROMIS Network (USA).
Discussants: Ms. Sally Goodspeed (Australia) and Ms. Jeannette Klimont (Austria).

Session 3: Development of Questions for each Domain
Presentations given by Canada, the USA, and Italy.
Discussant: Mr. Ian Mc Dowell (University of Ottawa, Canada)
.
Session 4: International Comparability
Presentations given by Eurostat and WHO.
Discussant: Mr. Gaetan La Fortune (OECD).

Session 5: Next Steps and Future Work
Presentations given by the USA, Netherlands, and Canada.
Discussant: Mr. Niels Rasmussen (EU Partnership Health/Denmark).

4. All papers and presentations from the meeting are available on the UNECE website: http://www.unece.org/stats/documents/2005.11.health.htm

SUMMARY OF THE DISCUSSIONS AND OF THE MAIN CONCLUSIONS REACHED AT THE MEETING

Session 1: Background, Terminology, and Scope

5. The meeting discussed the work of the UNECE Task Force on the measurement of health status. The meeting stressed the need for a clear expression of the purpose of the measure, its conceptual framework, and unambiguous definitions. The need to separate the description of the health state per se from its determinants and consequences was also stressed. There was general agreement on the need to measure health state preferably in terms of capacity, acknowledging that this might not always be feasible or appropriate (e.g. the pain/discomfort domain). The discussion recognized the need to situate health state into a framework for the measurement of health status, while acknowledging that other elements or domains are important.

6. It was noted that reaching international agreement in mapping survey questions to ICF concepts should be considered.

7. The TF expressed a clear preference for wording questions in terms of capacity. It was mentioned that the distinction between capacity and (actual) performance for measurement of health state, though conceptually clear, might not always be feasible for measurement purposes. The Eurostat approach has been slightly different in attempting to answer the question "Is individual A more disabled than individual B?" and has focussed on long standing difficulties in capacities but it also included difficulties in performance in the ADL and the IADLs. Individual choice and volition in part account for contrasts between capacity and performance.
How the data will be used and interpreted is an important determinant of how the measurement instrument should be developed. Data on health states may not be of much practical importance if systems for modifying policies based on the data are not in place. It was also emphasized that data generated through these types of instruments must be relevant for policy planning and be of practical significance. Complementary information may be provided by health examination surveys that ought to be also considered.

8. The questions of which domains to include and how the questions should be framed must be guided by empirical evidence. This would also hold true for asking the questions with or without assistance, aids (including medication), and adaptations. The need to co-ordinate the different internationally ongoing streams of work, including the work of Eurostat and the Washington Group in the development of a health status module, was stressed.

9. In the discussion the difficulties in reaching international comparisons were reiterated. The tensions between the need for continuity with ongoing data collection efforts in each country vs. the desire for comparability were recognized. Nonetheless, the fact that we need to move from outcomes that are essentially mortality based to non-fatal outcomes is obvious. The current exercise needs to build linkages to other ongoing efforts such as those of the PROMIS network for item banks for clinical populations, those of the Washington Group, and Euro HIS .

Session 2: Criteria for Selecting Domains and Identification of Domains

10. Based on the presentation of the criteria established by the Task Force, the discussants

focused on a number of issues relating to the selection of domains:

a) the need to evaluate through an exercise on how the criteria match with the selections made, such as by evaluating the relationships between health state measurement outcomes by socioeconomic status, ageing and other health determinants ;
b) the usefulness of the criteria for selecting the domains for which it is possible to obtain adequate data for a substantial part of the population.

11. Further points raised in the general discussion were:

a) the fact that the purpose of the measurement drives the selection of the domains as well as the items within the domain (e.g. emphasis needed on functioning/mobility of upper limbs for working age population);
b) to have rather more than less domains for which questions should be tested;
c) the possibility/need to obtain over time convergence on the domains selected through different processes : the Washington Group (WG), European Health Status Module (EHSM), UNECE/WHO/Eurostat meeting;
d) the possibilities for using the experiences obtained through patient-reported outcomes (e.g. the US PROMIS Network and the Dutch experiences) for population-based interview surveys;
e) whether or not 'pain' and 'vitality/fatigue' should be retained as separate domains for health state measurement;
f) patient-reported outcomes showing that some items – such as 'sexual dysfunctioning' and 'need for assistance' - 'cluster' outside the domains identified; and
g) the use of Item Response Theory (IRT) was proposed for selecting items for domains and the existence of item-banks (in US and NL) for clinical outcome measurement.

Session 3: Development of Questions for each Domain

12. The meeting reviewed the work carried out by the Task Force in developing core questions for each of the selected domains. Participants discussed general issues related to the design of the questions together with specific issues related to each domain. The experience of the U.S. PROMIS network enriched the discussion by providing useful information on their experience in selecting domains and designing questions. Data analysis and testing carried out by PROMIS could be a useful input to the work of the task force on finalizing the draft of the questions. The following issues were discussed on how to finalize a comprehensive and coherent set of questions to measure health state.

General Issues

a) Draw examples of questions related to the different domains not only from international/supranational modules (such as the WG and the European Health Status Module – EHSM), but also from the rich experience of national health interview surveys
b) Give explanations on how this set of questions relate to other processes such as the development of a disability module by the WG and the EHSM. Efforts could also be made in order to maximize the consistency between these three processes
c) Include criteria in the selection of domains and questions to make sure that the

identified questions can describe each domain with a small set of descriptive ordered categories

d) consider developing a preamble for the module and/or each domain/question
e) As the draft questions were developed in isolation for each of the selected domains, there is now a need to review the overall set in order to improve the coherence among the different domains and produce the module as one "package"
f) In order to assure the comparability of the definitions and the approaches used in measuring health states, instructions for interviewerers could also be developed
g) Clarify the total duration of the full module. Some indications were given for a 3-4 minutes module (n.b. – this deleted phrase was not in my notes, plus the number of questions is ambiguous, since sometimes there are multi-part questions)

13. The suggestion was also advanced to develop guidelines to standardize the design of questions in each health state domain These guidelines should however acknowledge that in some domains it may be not possible or reasonable to strictly follow them. The guidelines could cover the following:

a) The inclusion or not of the use of assistive devises and medications
b) The unidimensionality of questions. This is sometimes a tough criterion to apply and it should be viewed as a desirable approach to ensure clarity of the concept
c) The number of response categories. The PROMIS network has some empirical results that could help reach a standard number of categories; five was substantially suggested as optimal
d) The time frame. The PROMIS network has some empirical results to show the different reporting of respondent to different time frame
e) The use or not of filters
f) What aspect should be measured: intensity, frequency, duration, interference, quality, or location.

14. The PROMIS network has agreed to continue to share the protocols and other relevant materials as they become available. In addition, the UNECE TF items can be included in the PROMIS cognitive interviews, data collection (scheduled for April 2006) and data analyses. The groups will keep each other informed about the future developments and coordinate efforts when appropriate.

Mobility

a) Review the current wording together with the response categories
b) Under the goal of parsimony, consider reducing the number of items considered in this domain

Dexterity

a) In the total population dexterity problems have a low prevalence, but it can be a domain of higher relevance for older age groups. There could be the need to re-discuss the reasons and the relevance of including this domain in a parsimonious set of questions
b) Make the approach used and the use of specific objectives culturally relevant

c) Review the relevance of the two suggested questions; the second question could be replaced with a broader question dealing, for example, with reaching and related not only to the use of hands
d) Carefully discuss, the value added of a second question and its relevance, with the view of parsimony

Vitality/Fatigue

a) Tiredness is a difficult domain to measure since it appears as a cultural aspect more than a health aspect. Review the definition of this domain and consider using sleep duration and quality instead of tiredness
b) In view of parsimony the use of one question instead of two could be discussed. The testing of the EHSM could give the relevant information on this

Affect

a) Further analysis is needed for the questions related to happiness (positive affect). Between the two aspects of this domain, depression (negative affect) is more important when looking at decrements in health. However, the association between depression and happiness is open to different interpretations, and depression does not necessarily substitute for unhappiness
b) The two aspects could also both be negatively formulated in the questions.
c) The meeting also discussed whether anxiety and depression should be combined in a single domain as is done by existing instruments such as SF-36.

Anxiety

a) Remove references to medicines
b) In view of parsimony, review the need for different questions to measure intensity and frequency. Consider justifying the need to include both intensity and frequency and explore the possibility of combining intensity and frequency in one question with different response categories. For chronic conditions intensity could be a better measure while for other conditions frequency is a better measure
c) Consider making the language of the questions closer to the language spoken by the majority of people. Avoid academic and technical language. Wording of the type: "How often have you been worried …?" could be simpler

Pain

a) More analysis is needed in order to understand what is important to measure: intensity, frequency, and location? Location has been shown to be less important in accounting for the burden of pain, while frequency seems to be very important although the relevance of intensity and frequency change according to the type of condition.
b) All the questions deal only with pain and not discomfort
c) Although interference questions have proven to cover large parts of the pain domain, this is not the concept of "close to the skin" that is intended to be measured and it is

more a measure of impact

d) Consideration should be given to see if it is possible to combine intensity and frequency in one question

Hearing

a) It could be made more explicit why the difference between the conversation of one person or more than one person is made. If the purpose is to measure hearing in a situation with background noise where different people talk at the same time, review the questions
b) The introduction of scaled response categories could be considered
c) Further discussion is needed to consider the inclusion of hearing aids in the questions.

Vision

- Further consideration could be given on how to treat the use of glasses

Social Relationships

a) There was general agreement that the domain is important.
b) Further discussion is needed on the relevance and feasibility of including this domain. Conceptually it may be hard to translate the capacity of engaging in social relationship into the definition of health in different countries. Caution is advised when looking at the feasibility of measuring this capacity through HIS. It is not clear if an individual is able to report on his/her capacity to engage in social relationship
c) If this domain is included, consider narrowing down the concept that is intended to be measured and clearly define it
d) It is important to disentangle communication and relationships

Cognition

a) The aspect of memory seems to be the most difficult to measure. In particular with the current question it is hard to distinguish what a person thinks he/she is forgetting and what is the real capacity of a person to remember
b) It is also difficult to define a measure that works equally across different ages

Sessions 4 and 5: International Comparability, Next Steps and Future Work

15. Discussion focused on the issue of comparability both across subpopulations and time within populations and across countries, which cannot be ignored by national and international statistical agencies, and questions were raised about the quality of existing self-reported data. In this context, the need was highlighted for developing standardized well-designed survey instruments that are consistent across languages when using unanchored response categories.

16. It was acknowledged that an adequate conceptualization, elaboration/selection and subsequent translation of both questions and response categories cannot guarantee equivalent

responses from about similar populations in different countries. However, other strategies, such as the vignettes used in the WHO surveys and post-harmonization techniques, such as item response theory (IRT) were also presented. There was concern expressed about the comparability of the fixed levels described by the vignettes when translated into different languages. Some participants argued for the use of tests giving external calibrators rather than vignettes, which involve a second set of translation and self-reported issues.

17. There was considerable discussion on whether the vignette approach would remove valid cross-cultural differences in subjective experience of health states. An analysis of the self-reports in different domains of health in Sweden and Italy found that though the self-reported categorical results were very different between the two countries, the different results were primarily due to systematic differences in the way that the response categories were used. WHO responded that the vignettes were not addressing issues of perception or subjectivity in responses but rather the differential use of response categories for the same level of the item being measured (whether that item was objective, perceived or subjective).

18. It was reported that the vignettes are equally subject to measurement error as the self-reported of health, and will contain individual variations and biases. However, it was also argued that as long as the signal is not drowned by the noise, the vignettes provide useful information that can improve the comparability of self-reported health. There are examples where the vignettes allow the identification of substantial differences between groups that are otherwise hidden by the effects of cut point shifts.

19. It was shown that information theory can be used to select questions or entire questionnaires with the highest information contents, which may raise the efficiency of selecting domains/questions.

20. In order to ensure comparability among countries, the importance of ensuring translation of concepts of the survey modules in different countries was also emphasized. The experience of Eurostat on developed a translation protocol was presented. For ensuring comparability between the national versions of the Eurostat health survey, the protocol includes conceptual translation cards for each part of the module that have been prepared in order to help the translators to understand and translate the underlying health concepts.

21. The discussion noted the importance of including cognitive testing (also called debriefing), and detailed information on methodologies was presented. Using standard methodologies when implementing cognitive testing ensures greater levels of standardization across test sites and helps to consistently understand how the response mechanisms operate in the different countries.

Overall recommendations

22. For the next meeting, the documents should be prepared prior to the meeting date and made available to participants with adequate time for participants to review the documents.

23. For easier reference and identity purposes, it was recommended to label the current UNECE-WHO-Eurostat work on health as the Budapest Initiative.

FUTURE WORK

24. The meeting acknowledged the good and hard work carried out by the Task Force and the Steering Group and proposed that the two groups continue their work with their current memberships. Countries that would like to join one or both groups can communicate their availability to UNECE no later than two weeks from the end of the meeting.

25. Under the coordination of the Steering Group, the Task Force is asked to:

a) Build the relationship with other groups that are working on survey modules related to health state (in particular the Washington Group, WHO, and Eurostat)
b) Acknowledge and work with the US PROMIS (Patient Reported Outcomes Measurement Information System) network
c) Finalize Working Papers 1, 2, and 3, taking in consideration the suggestions provided by the meeting and summarized in the above-sessions of this report (March 2006), with a view to their eventual incorporation in the final document of the "Budapest Initiative"
d) Meet in March/April 2006 to finalize the selection of domains and draft set of questions and develop a testing plan that countries can implement
e) Coordinate and review testing done by member countries
f) Coordinate its work with Eurostat and take in consideration the results of the European testing as much as the timing of the analysis of these results will allow
g) To work toward the development of an handbook on the measurement of health state (March 2006-December 2007)

26. It is proposed that the next plenary meeting be organized in December 2006 with the objective of reviewing the work of the Task Force related to question development, review available testing results to date, modifying questions based on the discussion and review, and approve the revised draft set of questions.

27. It is proposed that the work in the next meeting be organized with some working group sessions in order to ensure the full participation of all the participants.

ADOPTION OF THE REPORT

28. The participants adopted the report of the meeting at its closing session.
