

Working Paper No.3
1 November 2005

ENGLISH ONLY

**STATISTICAL COMMISSION and
UN ECONOMIC COMMISSION FOR
EUROPE**

**STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES
(EUROSTAT)**

**CONFERENCE OF EUROPEAN
STATISTICIANS**

**WORLD HEALTH
ORGANIZATION (WHO)**

Joint UNECE/WHO/Eurostat Meeting
on the Measurement of Health Status
(Budapest, Hungary, 14-16 November 2005)

Session 3 – Invited paper

**CONCEPTUAL AND LOGISTIC ISSUES IN ITEM CONSTRUCTION AND PROPOSED
QUESTIONS FOR DOMAINS. ***

Submitted by the Task Force on Health Status

Background

1. Building on the agreements reached at the joint UNECE/WHO/Eurostat Meeting on the Measurement of Health Status held in Geneva, May 24-26, 2004, the Task Force (TF) on the development of a common instrument for measuring health states has (1) established a guiding conceptual framework regarding health states and their measurement, and (2) used pre-specified criteria to systematically reduce an initial list of 30 candidate health domains into a more parsimonious set of 10 domains for inclusion on the common instrument.¹ The selected domains are as follows, and will be assessed in terms of an individual's *capacity* or *ability* to function:

- 1) Physical Functioning: Mobility
- 2) Physical Functioning: Dexterity
- 3) Vitality/Fatigue
- 4) Affect
- 5) Anxiety
- 6) Vision
- 7) Hearing
- 8) Pain and Discomfort

* This paper is Part 3 of 3 papers prepared by the Task Force on Health Status for this meeting. The Task Force consists of representatives from the following countries and international organizations: Australia, Belgium, Canada, Estonia, Hungary, Italy, Netherlands, Norway, Spain, United Kingdom, USA, Eurostat, UNECE, and WHO.

¹ Task Force on the Measurement of Health Status. Part 1 of 3 – *Health as a multi-dimensional construct and cross-population comparability*.

- 9) Social Relationships (including aspects of communication)
- 10) Cognition
 - a. memory and concentration
 - b. problem solving and thinking

2. The TF has also developed the next phase of the project, namely the drafting of survey items for measuring each of the above domains. The following general issues were raised regarding the form of the common instrument: (1) the number of survey questions necessary to tap each domain, (2) ensuring the uni-dimensionality of the items, (3) the duration of the recall period provided to respondents, (4) how to address the effects of both technical and medical prosthetics (e.g., glasses, walking equipment, painkillers, antidepressants, etc.), and (5) use of meaningful response categories amenable to both translation and preference measurement. Since these issues pertain to question design for all domains, they are discussed here at a general level and some overall guidelines are provided. Although no “golden rules” for item construction have been discovered, efforts have been made to narrow the range of options for potential cognitive field testing.

Survey Item Issues

I. Number of Questions per Domain

3. Given that the common instrument will likely form one component of a broader health survey instrument (administered in person or by telephone), care should be taken to ensure that each domain has enough items to measure it in sufficient depth for obtaining the desired data, but not so many questions that the operational feasibility of the survey is compromised. Since the objective is the generic assessment of population health status rather than a detailed clinical evaluation of persons, a small set of broad questions, which are in line with the guiding conceptual framework of the TF, should be able to provide adequate coverage of the domains of interest. Generic instruments such as the Health Utilities Index Mark 3 (HUI3) have enjoyed considerable success in assessing population health status on national surveys with a compact yet comprehensive set of questions.^{2,3} Reviewers in the more general fields of subjective well-being and quality of life suggest that if broad assessments are the objective of the study, then single item scales might be able to produce sufficient information on a domain.^{4,5}

4. Ideally, then, the number of questions for each domain on the common instrument should be limited to one or two, as was done on the World Health Survey instrument.⁶ However, with respect to the TF’s criterion of item uni-dimensionality (discussed in more detail below), it is important to point out that more questions may be necessary in the case where domains have multiple facets, which cannot be meaningfully collapsed into a single question. A key example of such a case would be the Cognition domain, which includes the sub-domains of memory,

² Feeny DH. Health status classification systems for summary measures of population health. In: Murray CJL, Salomon JA, Mathers CD, Lopez AD, editors. *Summary measures of population health: Concepts, ethics, measurement and applications*. Geneva: World Health Organization; 2002. p. 329-341.

³ Feeny DH. The Health Utilities Index: A tool for assessing health benefits. QoL Newsletter 2005; 34:2-6.

⁴ Diener, E. Subjective well-being. *Psychological Bulletin* 1984; 95: 542-575.

⁵ Diener E, Sandvik E, Pavot W, Gallagher D. Subjective well-being: three decades of progress. *Psychological Bulletin* 1999; 125: 276-302.

⁶ World Health Survey (WHS): New Instrument and Related Documents.

<http://www3.who.int/whs/P/instrumentandrel8293.html>. Accessed July 2005.

concentration, thinking, and problem-solving. Thus, it may be necessary to have up to four questions in order to deal with this and other multi-faceted health domains.

5. It should be noted, however, that the above recommendations are not intended to minimize the development of items in the early phases of the TF's work, and should therefore not hinder the creation of an initially more extensive item pool. At this stage, even if more face-validated items than are feasible to place on the common instrument are identified, they have been included for review and consideration by the TF. Once all of the items are pooled, a collaborative and iterative process involving polishing, trimming, and rearranging will then be conducted, in order to arrive at a suitable subset of items for the common instrument. Further, any versions of the common instrument used for cognitive pre-testing could include more items than will be used in the final tool. Respondent feedback could then be used to inform additional cutting and other refinements.

6. *Filter questions.* Another issue under the more general considerations of number of items and operational feasibility is the use of filter questions. Filter questions could reduce cognitive burden and conserve survey time. Specifically, by posing an initial question determining whether the respondent had any functional limitations whatsoever on the domain in question, it would not be necessary in certain cases to read all of the response options. For example, filter questions could take the following form:

During the past ___ days, did you have any difficulty walking around the neighbourhood?
Yes
No

During the past ___ days, did you have any pain or discomfort?
Yes
No

Negative answers could simply be coded as normal functioning, while the follow-up questions for affirmative responses could be, for instance:

During the past ___ days, how much difficulty did you have in walking around the neighbourhood?

- 1) a little bit of difficulty
- 2) quite a bit of difficulty
- 3) a great deal of difficulty
- 4) complete difficulty/unable to do it at all

During the past ___ days, how would you describe the overall intensity of your pain or discomfort?

- 1) mild
- 2) moderate
- 3) severe
- 4) extreme

7. Given that normal functioning on the selected domains will likely characterize a large proportion of respondents on a general population survey, filter questions may enhance efficiency. However, there is the possibility that some respondents may elect to indicate no

functional limitations if they perceive they can reduce the interview time in this fashion. Schaeffer and Presser⁷ point out that full filter questions offer participants relief from the cognitive effort necessary to construct a response. Thus, one potential trade-off with using filter questions is that we may miss some of the burden – respondents having functional limitations could systematically provide negative responses to filter questions such as the above. Perhaps more importantly, questions with only two response categories (e.g., yes versus no) could have different operational characteristics than questions offering a greater range of response categories. Even if the filter question does not discourage certain participants from thoughtful consideration of their true level of function, those with milder limitations on a given domain might still answer “no” to the question with just two options, yet indicate the presence of limitations if the question provides them with response options permitting finer differentiation (e.g., mild, moderate, severe, etc.).

8. Overall, using filter questions does have some operational advantages, but in light of the above considerations, it might produce a very different distribution of functioning than simply presenting items with a full suite of response categories. Cognitive testing would therefore be needed to determine whether filter questions introduce any type of response bias. For example, it could be examined whether the frequency of normal levels of functioning is significantly higher in groups administered filter questions than in those directly asked to describe their domain-related functioning (i.e., where all possible functional levels are presented, in whatever form, to respondents).

II. Questions Should Be Uni-Dimensional

9. There is agreement among TF members that each question on the common instrument should reflect one concept only. In Paper 1, which initially outlined the overarching conceptual framework of the TF, it was recommended that efforts be made toward generating *uni-dimensional* items.¹ Indeed, it can be problematic when numerous conceptually and empirically distinct, yet related aspects of functioning are subsumed under a single survey item. A single item, which mixes multiple concepts is typically referred to by survey methodologists as “double-barreled,”⁸ as it essentially requires respondents to provide a single answer to multiple questions. For example, the EQ-5D instrument⁹ contains a domain, which combines anxiety and depression; and respondents are asked to indicate which of the following three levels is most descriptive of their current emotional state:

- 1) I am not anxious or depressed
- 2) I am moderately anxious or depressed
- 3) I am extremely anxious or depressed

10. Answers to such questions might be problematic to interpret, since it cannot be determined precisely which part of the question the respondent was answering. While anxiety and depression are frequently co-morbid with each other¹⁰, they are both conceptually and empirically

⁷ Schaeffer NC, Presser S. The science of asking questions. *Annual Review of Sociology* 2003; 29: 65-88.

⁸ Ellard JH, Rogers TB. Teaching questionnaire construction effectively: The Ten Commandments of question writing. *Contemporary Social Psychology* 1993; 17: 17-20.

⁹ Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQoL group. *Ann Med* 2001; 33: 337-343.

¹⁰ Kasper S. Depression and anxiety –separate or continuum? *World J Biol Psychiatry* 2001; 2:162-163.

distinct domains of mental health.^{11,12} Persons who are anxious but not depressed, or depressed but not anxious, may have a difficult time responding to such a question. Another prime example is the case of the Cognition domain, multiple facets of which are often assessed by a single question on health surveys and generic health status instruments. If a single item refers to functioning on multiple facets of Cognition (e.g., memory, concentration, problem-solving, and thinking), yet permits only one global rating of functioning (e.g., ranging from normal to extremely limited functioning), it may create ambiguity for survey respondents, as well as render it problematic to determine precisely where people's reported cognitive deficits lie.

11. Therefore, constructing uni-dimensional questions can potentially improve the clarity of items for respondents and provide further refinement in the assessment of health states. In addition, this strategy could also render any future preference elicitation exercises less mentally burdensome for raters, since each level within the health state description system would describe functional limitations pertaining to only one concept. It is important to note, however, that efforts toward achieving uni-dimensionality of questions and refinement of domain measurement should not compromise the operational feasibility of the common instrument. In order to remain in line with the above criterion of having a workable number of questions per domain, it may be necessary to select a manageable subset of the most important concepts related to each domain.

III. Duration of the Recall Period for the Questions

12. Specification of a recall or reference period in the items defines the particular time span respondents should use for estimating their level of functioning on each domain. As discussed in more detail below, the major advantages in using a specific recall period are that it would improve the reliability and comparability of responses, as well as facilitate translation into numerous languages. At present, however, there is no consensus regarding the inclusion of a recall period in the common instrument. Given that a time reference could probably be incorporated into the questions without fundamental changes to content, it has been agreed to reserve the issue until the later stages of the work. Nonetheless, the current paper addresses in more detail the use of particular recall periods with survey questions, in order to facilitate future discussions on potential options for the common instrument.

13. First, the specification of a recall period depends largely on the objectives of the research. For example, in order to assess health status in clinical studies and economic evaluations where researchers are interested in monitoring transitions, the HUI3 uses three standard "current" recall periods: one, two, and four weeks.¹³ If the situation is one in which fairly rapid changes in health status are believed to be occurring (such as acute illness onset and recovery, minor surgery), the one- and two-week recall periods may be most suitable. On the other hand, in the context of chronic disease or recuperation from major surgery, the 4-week time span may be better to use. The standard alternative version of the HUI3 is the "usual" version, which does not define a specific recall period. Rather, the questions ask about usual abilities and feelings; for example:

¹¹ Endler NS, Denisoff E, Rutherford A. Anxiety and depression: evidence for the differentiation of commonly co-occurring constructs. *Journal of Psychopathology and Behavioral Assessment* 1998; 20: 149-171.

¹² Mykletun A, Stordal E, Dahl AA. Hospital Anxiety and Depression (HAD) scale: factor structure, item analyses and internal consistency in a large population. *British Journal of Psychiatry* 2001; 179: 540-544.

¹³ Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI®): concepts, measurement properties and applications. *Health and Quality of Life Outcomes* 2003, 1:54

How would you describe your usual ability to think and solve day-to-day problems? Are you usually free of pain or discomfort? The “usual” version often employed in population health surveys, in order to avoid picking up the burdens of time-limited illnesses and conditions¹⁴; a version that specifies functioning over the past 3 months has also been used. Further, the “usual” version might avoid confounding a person’s intrinsic capacities with limitations in functioning resulting from recent changes in environmental factors. However, some cognitive testing has suggested that not all respondents interpret the term “usual” in the anticipated manner.

14. In addition, it could be problematic to translate “usual” into a wide variety of languages, as will be required for the TF’s common instrument. A specific recall period would be more easily translated, since it involves units of time, which are not as subject to multiple interpretations and other culture-specific biases. Further, if only the terms “usually” or “generally” are used to set the context, it seems likely that there will be high within-respondent variability regarding the length of the recall period used to derive an estimate of functioning on a given domain. For a potentially wide variety of reasons, some participants may select the prior month to construct a response, while others may choose to review the previous year of their lives. Therefore, standardizing the recall period across respondents might minimize the influence of such idiosyncratic factors. However, there are a number of issues to be considered when choosing a specific recall period for survey items. The next few paragraphs briefly review some research findings on this issue, which may be informative regarding the eventual design of the common instrument.

15. A number of studies have shown that the duration of the recall period influences the types of events that survey respondents take into account. For instance, in one experiment, Winkielman, Knäuper, and Schwarz¹⁵ varied the recall period for reporting the frequency of anger episodes: the number of episodes experienced over the past week versus those occurring in the past year. It was found that respondents in the one-week condition thought that the researcher was probing for frequent, low-intensity episodes of anger; whereas in the group asked to give a 1-year retrospective report, it was inferred that the researcher was interested in more rare and severe episodes of anger. Other studies on the current versus retrospective reporting of emotions have produced similar findings;^{16,17} also, couples tend to report more severe marital disagreements when asked about their past versus current situation.¹⁸ In consideration of this work, it appears that people’s reports across different recall periods may be incomparable, even if the same basic information is being requested.¹⁹ Therefore, if a reference period is to be used for the TF’s common instrument, it may be desirable to try and specify one that is not so long that respondents

¹⁴Swain, L., Catlin, G., & Beaudet, M.P. (1999). The National Population Health Survey – its longitudinal nature. *Health Reports*, 10(4), 69-82.

¹⁵ Winkielman P, Knäuper B, Schwarz N. Looking back at anger: Reference periods change the interpretation of (emotion) frequency questions. *Journal of Personality and Social Psychology* 1998; 75: 719-728.

¹⁶ Parkinson, B, Briner RB, Reynolds S, Totterdell P. Time frames for mood: Relation between momentary and generalized feelings of affect. *Personality and Social Psychology Bulletin* 1995; 21: 331-339.

¹⁷ Thomas DL, Diener E. Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology* 1990; 59: 291-297.

¹⁸ McGonagle KA, Kessler RC, Schilling EA. The frequency and determinants of marital disagreements in a community sample. *Journal of Personal & Social Relationships* 1990; 59: 291-297.

¹⁹ Schwarz N. Self-reports: How the questions shape the answers. *American Psychologist* 1999; 54(2): 93-105.

only focus on the most intense events related to the domains of functioning, but not so short that only highly frequent and patterned events of lower intensity are considered.

16. Another noteworthy research finding is that events, which fall outside of a specified recall period can still influence responses to survey items; specifically, this phenomenon is referred to as *telescoping*.⁷ Events that are more vivid and intense are more likely to be readily available in memory, so even if they are not strictly within the recall period, they may be included due to participants' use of this "availability" heuristic. More precisely, improper inclusion of an event in a recall period is called *forward telescoping*. With events that are less intense and not as well-represented in memory, *backward telescoping* may occur, that is, an event can be improperly excluded from the recall period. Thus, in order to try and enhance measurement precision, the preamble could include a statement such as:

"When answering these questions, please try and focus only on your life situation during the past ___ days."

17. In sum, there are numerous complex factors to consider when deciding on a specific recall period for the common instrument. The most basic necessities would seem to be that the selected recall period should appear in the preamble, be identical across all questions, and appear in the same location in all questions. For example:

"The next set of questions deals with your abilities and feelings during the past ___ days. When answering these questions, please try and focus only on your life situation during the past ___ days".

1. *During the past ___ days, how much difficulty have you had in doing ___?*

2. *During the past ___ days, how much of a problem did you have with feeling/being ___?*

18. Repeating the recall period reduces the probability that respondents will lapse into different personal recall periods, and keeping the location of the recall period constant across questions helps ensure that each question is processed in a parallel manner (i.e., that the respondent does not determine his or her own recall period before the interviewer reads it).⁷

19. Some of the issues described above could be addressed in cognitive pre-testing in which the recall period was systematically altered (e.g., none, one week, two weeks, one month, two months). Differences among experimental groups in terms of self-reported functional status could then be assessed, and the extent to which the recall period accounted for the observed variation could be estimated. Further, participants in each group could be probed about the types of situations they were thinking about when answering questions on the ability and "feeling state" domains, in order to see if their answers may have been unduly levered by considering a preponderance of certain types of events (e.g., a focus on times when they had severe versus mild difficulty in executing tasks or actions, or a focus on episodes of high- versus low-intensity feelings – affect, anxiety, and pain).

IV. Dealing with Technical and Medicinal Prosthetics

20. Since the objective of the current exercise is to measure peoples' basic capacities and thus adhere as much as possible to a "within-the-skin" concept of health states, questions for particular domains may have to include information on prosthetics in order to discern what respondents are capable of doing without the use of aides. Otherwise, certain cases may be incorrectly recorded as

having no functional limitations, simply because the question did not provide any information on the potential role of prosthetics in executing the task or action of interest. For example, a question created to assess basic mobility limitations might be worded as follows, with no information included on prosthetics: **“During the last __ days, how much difficulty did you have in getting around the neighbourhood?”** Given the form of this question, both individuals with and without full use of their legs could conceivably report having no difficulty, provided that the latter had mechanical aides (e.g., a walker or wheelchair). Hence, the same level of functional capacity could be assigned across persons irrespective of inter-individual differences in the true level, resulting in measurement error. Further, if no mechanical aides are mentioned in such questions, it is possible that some interviewers may receive questions to the effect of: **“Do you mean, can I get around the neighbourhood with or without my... walker/wheelchair?”** In order to overcome such measurement dilemmas, the HUI3 Ambulation domain explicitly incorporates information about walking equipment, as well as the degree of assistance needed from other persons¹³ :

- 1) Able to walk around the neighbourhood without difficulty, and without walking equipment.
- 2) Able to walk around the neighbourhood with difficulty; but does not require walking equipment or the help of another person.
- 3) Able to walk around the neighbourhood with walking equipment, but without the help of another person.
- 4) Able to walk only short distances with walking equipment, and requires a wheelchair to get around the neighbourhood.
- 5) Unable to walk alone, even with walking equipment. Able to walk short distances with the help of another person, and requires a wheelchair to get around the neighbourhood.
- 6) Cannot walk at all.

21. While including a clause that the respondent is to reply assuming that assistive devices are not used should clarify the meaning of the question and improve comparability, some cognitive testing has shown that in some cases, the respondent interprets questions that mention assistive devices as questions about the device rather than about their functioning without the device, that is, they just answer that they don't use the device (*references forthcoming*). Further, asking that the respondent report difficulties without the use of aids might also be problematic for aids that are widely used and very effective, such as glasses. Questions on the degree of visual acuity without glasses would identify a large segment of the population in many countries, and this group would include many persons with very mild limitations. While it is recommended that the questions on the TF's common instrument relating to physical functional domains should include information on prosthetics and assistive devices, this should only be done where relevant and where it will permit better assessment of the true extent of the underlying functional limitations.

22. The above issues regarding prosthetics will also apply to the items used to measure the “feeling state” domains (i.e., Pain and Discomfort, Anxiety and Affect). In this case, the use of medicinal aides is an important consideration in item construction. Some health status instruments explicitly provide information about medication in order to help gauge the severity of the feeling state of interest, such as the pain domain on the HUI2:¹³

- 1) Free of pain and discomfort.
- 2) Occasional pain. Discomfort relieved by non-prescription drugs or self-control activity without disruption of normal activities.
- 3) Frequent pain. Discomfort relieved by oral medicines with occasional disruption of

normal activities.

- 4) Frequent pain; frequent disruption of normal activities. Discomfort requires prescription narcotics for relief.
- 5) Severe pain. Pain not relieved by drugs and constantly disrupts normal activities.

23. It should be noted that the HUI2 was designed specifically for assessing the health status of children with cancer. Indeed, if one wishes to link functional status to specific health conditions, information on medications can provide valuable information. By contrast, the HUI3 was created largely for use on general population surveys (although it is often used in clinical studies²⁰) and focuses exclusively on the degree of activity disruption as the index of pain severity. The use of medication is not mentioned in any of the HUI3 Pain items. The TF's common instrument will also be oriented toward assessing the health of general populations, and therefore one option might be to draw on the HUI3-type approach when constructing items for its feeling state domains (i.e., Pain and Discomfort, Anxiety, Affect).

24. However, if some respondents who use medication tend to base their answers primarily on their medicated states, while others who use medication tend to give more weight to their non-medicated states, some measurement inaccuracies and inconsistencies could occur in a population survey. If information on medication does not appear anywhere in the preamble or the question, some respondents may wonder if the interviewer wants to know about their medicated or non-medicated states. Interviewers would therefore have to be prepared for a query such as: **“Are you referring to the intensity of my Pain and Discomfort while I am on my medication, or not?”** Such a question may be quite natural for a person who has to manage chronic pain (e.g., from arthritis).

25. One approach to addressing the impact of medication would be to try and standardize interpretation of the feeling state items across respondents. For example, on the Pain and Discomfort attribute, a statement such as the following could be added to the preamble:

“When answering this question, please consider any medication you are using and the degree to which it relieved your Pain and Discomfort over the last __ days.”

26. With such a statement included, the survey question itself would not need to explicitly provide information on medication, thereby conserving survey time. Similar approaches could be used for the Anxiety and Affect domains, and the selected statements could be pre-tested for comprehensibility and effectiveness.

V. Item Wording and Response Categories

27. Since the common instrument will need to be adequately translated into numerous different languages, it is important to choose the terminology carefully in order to facilitate this process. First, as mentioned above, the terms “usual” or “general” for setting the question context seem to be less amenable to translation than a specific recall period; in other words, units of time will probably have more consistent meaning across different languages and cultures than would the words “usual” or “general.” Second, in order to make the survey concepts as understandable for the general population as possible, the phrase “difficulty in doing ____” should probably be retained as the terminology of choice for the questions on the physical ability domains. Although *capacity* is actually the construct of interest, it probably cannot be tapped directly but rather only

²⁰ Furlong WJ, Feeny DH, Torrance, GW et al. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med* 2001; 33: 375-384.

indirectly through assessing the degree of difficulty people have in meeting the various demands of daily living. Respondents can probably report more easily on the extent of difficulty they have in executing certain daily life tasks or actions than they can on their underlying capacity or ability; the latter would seem to be more of a latent construct. So in this way, a scientific or *analytical construct* (i.e., capacity) is being mapped to a more lay-oriented or *native construct* (i.e., “difficulty in doing”), which is usually necessary in survey research for measuring the construct of interest.⁷ This was the approach generally taken on the World Health Survey (WHS);⁶ for example, for the mobility domain: **“Overall in the last 30 days, how much difficulty did you have with moving around?”** In addition, it seems likely that the approach of using “difficulty in doing ___” would pose less problems for translation than employing the somewhat more scientific and abstract jargon of “capacity for doing ___” or “ability to do___.”

28. As for selecting particular adjectival or adverbial qualifiers for the items, there are numerous issues and options to consider. Above all, from a purely operational point of view, a balance is needed between covering health status in adequate depth and keeping the survey within time limits. Therefore, similar to the criteria regarding the total number of survey items, there should not be so few response options that the categorization of functional limitations is too crude, yet not so many that it becomes difficult for interviewers to read them all and for respondents to cognitively manage the information.

29. Consideration of prior work could inform the selection of the optimal number of response options for the TF’s common instrument. A simple yes/no response format is probably too coarse to adequately tap a domain. The EQ-5D, which uses a 3-level descriptive system (e.g., no, moderate, and extreme problems), has been widely employed on population surveys. However, some investigators have needed to modify the EQ-5D in order to increase its coverage and sensitivity. In particular, for use in health state valuation exercises, *five* rather than three levels per domain were used in one study.²¹ Indeed, using five response categories (i.e., levels of functional limitations) for the TF’s common instrument would seem to be quite reasonable as well, based on some classic psychological research in the areas of human information processing and questionnaire development. First, empirical evidence suggests that five to nine independent pieces of information represent a maximal cognitive load for most persons.^{2,22} Second, another classic study on scaling for attitude questionnaires showed that five response categories appears to be the most efficient, and that increasing the number of categories beyond five resulted in a progressive loss of discriminative power.²³ More specifically, with five response categories, the highest degree of overlap in response distributions for any two adjacent scale points was 2%; with nine categories, it was 44%.

30. Additional, more challenging issues than the sheer number of response categories include the selection of the precise adverbial or adjectival qualifiers that should be attached to them, as well as identifying what aspect of “difficulty in doing ___” is being measured for the domain in question. More precisely, two aspects of “difficulty in doing ___” are candidates for study: level of

²¹ Mahapatra P, Nanda L, Rajshree KT. The 6D5L description system for health state valuation. In Murray CJL, Salomon JA, Mathers CD, Lopez AD. (eds.), *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*. Geneva: World Health Organization, 2002: 349-367.

²² Pearmain D, Swanson J, Kroes E, Bradley M. *Stated preference techniques: a guide to practice*. Hague, Davis Gleave and Hague Consulting Group; 1993.

²³ Bass BM, Cascio WF, O’Connor EJ. Magnitude estimations of expressions of frequency and amount. *Journal of Applied Psychology* 1974; 59(3): 313-329.

difficulty and how often the difficulty occurs. The specific qualifiers then serve to define the gradations in severity concerning the functional limitations in question. For example, some commonly used qualifiers are “none, mild, moderate, severe, and extreme/cannot do.” This set of qualifiers is widely used throughout the *International Classification of Functioning, Disability and Health (ICF)*²⁴ to describe varying degrees of impairment in Bodily Functions and Structures, varying degrees of difficulty with Activities and Participation, and variance in the potency of Environmental Barriers. The SF-36 instrument²⁵ uses frequency qualifiers on certain items to define the extent of functional limitations, for example: “all of the time, most of the time, some of the time, a little of the time, and none of the time.” Indeed, a review of a broad range of health status instruments reveals that there are too many rather than too few options concerning response category qualifiers.²⁶

31. With such a vast assortment of qualifiers available, the key objective is to choose a set, which is most amenable to translation, that is, one which has the most consistent meaning across countries and cultures. As discussed in more detail in Paper 1, international surveys in particular are faced with a “response category cut-point shift” problem.¹ More specifically, with regard to severity of functional limitations, a given limitation could be viewed as “mild” within one culture and “severe” within another. Similarly, with regard to frequency, notions of what constitutes experiencing a particular problem “a little of the time” versus “some of the time” could differ dramatically among cultures. Further, attempting to be more objective by using absolute frequencies (e.g, 10%, 20%, 30%) rather than subjective or relative frequencies (e.g., “rarely, sometimes, often”) may not get around this problem entirely, for a couple of reasons. First, it could impose a measurement system, which may not be compatible with a wide variety of cultural perceptions and preferences. Second, as Schwarz¹⁹ notes, survey respondents might still mentally convert absolute to relative frequencies anyway (and vice versa), prior to answering questions. Overall, concerning the potential use of intensity qualifiers and frequency quantifiers, it may be beneficial to begin systematically searching for cross-culturally equivalent terms, with the assistance of translators. The results could then be pooled and built into sets of questions for cognitive testing.

32. An additional issue that should be raised here is whether the *same* set of qualifiers will ultimately be used for both types of domains, namely the domains related to mental and physical capabilities and the “feeling state” domains. It would probably be easier on respondents if the same set of qualifiers were used throughout, since a shift in qualifiers might disrupt the flow of information processing to some degree. Of course, using the same set of qualifiers for the ability and “feeling state” domains would require that the questions for the latter be worded accordingly. For instance, if the qualifiers “none, mild, moderate, severe, and extreme/cannot do” were used throughout all questions, a set of questions on the Mobility and Pain and Discomfort domains might be structured as follows (similar to the approach taken on the WHS):⁶

1. *During the past ___ days, how much difficulty have you had in walking around the neighbourhood ___ ?*

2. *During the past ___ days, how would you describe your overall level of Pain and Discomfort ___?*

²⁴ World Health Organization. *International classification of functioning, disability, and health*. Geneva: World Health Organization; 2001.

²⁵ Ware JE, Jr. SF-36 health survey update. *Spine* 2000; 25(24): 3130-3139.

²⁶ Salek S (editor). *Compendium of Quality of Life Instruments*. Chichester, UK: John Wiley & Sons; 1998.

33. *Alternatives to full suites of response options.* Some specific alternatives to using sets of finely differentiated response options should be mentioned, for discussion purposes. For example, one possible way to minimize some problems in translation and cross-cultural equivalency is to label only the endpoints of a numeric scale, (e.g., 1 = “no difficulty” 2 3 4 5 = “extreme difficulty/cannot do”). Similarly, a *visual analogue scale* (VAS) approach may be worthwhile to consider for particular domains, especially the feeling state domains. Typically, the VAS is used to rank complete health states in valuation studies,²⁷ but could be applied to individual health domains as well. For example, one could use the following VAS for measuring Pain intensity:

2. *During the past ___ days, how would you describe your overall level of Pain and Discomfort ___? Please use the following scale:*



An item applying the VAS strategy to an ability domain (mobility) might be designed as follows:

1. *During the past ___ days, how much difficulty have you had in walking around the neighbourhood ___? Please use the following scale:*



34. While the description of the scale endpoints would still require translation, it would be substantially easier than in the case where all scale values were associated with written descriptions. Such a scale might also take the form of a ruler or ladder with descriptors on the endpoints. One widely used measure of subjective well-being is Cantril's *Life Ladder*,^{28,29} where the top (10) and bottom (0) rungs are described, respectively, as the best and worst possible lives imaginable. One potential advantage to using a ladder is that it would likely represent a more cross-culturally recognizable and comparable symbol.

35. However, as Schaeffer and Presser⁷ point out, measurement precision can be compromised when descriptors are not attached to all scale values. Further, for the purposes of preference measurement, it may be optimal to have a health state descriptive system where every domain level is clearly defined. Using vectors of clearly defined domain levels is a widely accepted method for measuring health state preferences, especially in lay populations.¹³ In order to develop a reliable and valid preference-based scoring function for use with the data ultimately obtained using the common instrument, the health state descriptions would require an explicit link to the items used. In order to satisfy this criterion, it may be best to explicitly define all domain levels. On the other hand, if maximizing cross-cultural comparability is deemed the most

²⁷ Torrance GW, Feeny DH, Furlong W. Visual analog scales: Do they have a role in the measurement of preferences for health states? *Med Decis Making* 2001; 21: 329-334.

²⁸ Cantril H. *The pattern of human concerns*. New Brunswick, NJ: Rutgers University Press; 1965.

²⁹ Kilpatrick FP, Cantril H. Self-anchoring scaling: a measure of individuals' unique reality worlds. *Journal of Individual Psychology* 1960; 16: 158-173.

important criterion, then a simplified VAS-type format might be helpful. Given these inevitable trade-offs, it is recommended that both types of items should be subjected to cognitive testing: items with verbal descriptors linked to all scale points, and items using scales where only the endpoints are labeled. For example, respondents could be asked about the ease with which they were able to locate their own level of limitation on each type of scale, as well as the perceived accuracy of their answer.

36. Overall, there are several issues that must be considered when constructing international surveys and striving for measurement equivalence.^{30,31} It is hoped that the above material can stimulate further discussion and debate among TF members as the work continues.

³⁰ Harkness JA, Schoua-Glusberg A. Questionnaires in Translation. *ZUMA-Nachrichten Spezial* 1998; January: 87-126.

³¹ Johnson TP. Approaches to equivalence in cross-cultural and cross-national survey research. *ZUMA-Nachrichten Spezial* 1998; January: 1-40.

Domains

1. Physical Functioning: Mobility
2. Physical Functioning: Dexterity
3. Vitality/Fatigue
4. Affect
5. Anxiety
6. Vision
7. Hearing
8. Pain and Discomfort
9. Social Relationships (including aspects of communication)
10. Cognition
 - a) memory and concentration
 - b) problem solving and thinking

Domain 1. Physical Functioning: Mobility

37. Questions on physical functioning, especially mobility, are found on most health surveys. While mobility can refer to movement from one place to another using any method, the definition of health state being used requires that the movement be accomplished by the individual without the use of assistive devices such as wheel chairs. Mobility is most often characterized by walking and climbing steps. Bending, stooping, and crouching, while not directly related to mobility, characterize another important aspect of lower body function and are often included in indicator sets that also contain walking and climbing steps. While climbing steps and bending, crouching and stooping are important aspects of lower body functioning, walking by far is the most important.

38. Difficulty walking is often related to musculoskeletal conditions, it can also be related to difficulty seeing, conditions of the heart, COPD and other illnesses. As a result, walking is not only the most central indicator of within the mobility domain; it is probably one of the most central indicators across all domains. When multiple indicators covering a range of domains are used to identify a group of individuals with difficulty on any one indicator, more of these indicators have difficulty walking than have difficulty with any of the other indicators. In some analyses the overlap is quite high. A decision must be made as to whether walking difficulties should be limited to those related to lower body function. Not including walking difficulties that relate to vision problems would satisfy the criteria of independence of domains while not including walking difficulties that are due to heart disease would violate the criteria of being agnostic with regard to etiology.

39. Questions from the European Health Interview & Health Examination Surveys Database were reviewed to identify candidate questions on mobility (the full list of questions is available for review). A review of these questions highlighted differences in the ways that these questions have been asked.

40. Description of the activity: Some questions ask directly about the activity, such as walking or climbing steps. Others ask more general questions referring to getting around (inside and outside), moving or moving around but the intent is probably to obtain information about walking.

41. Use of personal assistance or assistive devices: Whether or not to specify that the activity is performed with or without the use of assistive devices is an important consideration for mobility questions. There is a lot of variability in questions with some specifying that the activity is to be done unassisted, some specifying that the activity be done with help or the use of assistance and some not specifying whether or not the response should be based on assisted or unassisted performance. The question of whether and how to incorporate assistive devices has been discussed as a general issues applying across domains. To be consistent with the definition of health state, mobility should be evaluated without the use of assistive devices. However, some cognitive testing has shown that while the inclusion of phrases such as “without the use of devices such as wheelchairs or walkers” should improve the quality of responses, they often confuse rather than clarify the question. In addition to making the question wordier, subjects sometimes interpret the question as being about device use. They might have difficulty but since they don't use the device, they do not report the difficulty.

42. Specificity of the question: Mobility questions vary in the level of specificity with which the activity is described. The question can mention walking in general or can specify how long and/or how fast the walk is. Following general best practices in survey design, the more specific

the question, the more consistent and accurate the response. However, for this to occur, the specification must be understood by the respondent. Question wordings range from specifying walking a few steps to walking long distances or running. Some questions specify that the activity needs to be done without stopping or resting. Decisions must be made about selecting the specification. Asking about a few steps obtains information on severe mobility limitations. Extending the distance to capture less severe difficulties runs into problems of selecting the right distances and describing them appropriately. Do respondents understand how long a kilometer is?

43. Four questions are suggested for this domain – three covering walking/running and one covering climbing steps. The questions refer to the activity itself not moving around. No attempt is made to limit the questions to lower body related difficulties in performing the activities. While this does result in some overlap among the domain, including the necessary instructions in the question would be difficult.

44. The three walking questions are intended to tap different levels of severity. The actual distances mentioned are for illustration only but the idea is to capture severe difficulty as well as difficulty at longer distances. Skip patterns can be used for these questions. The question on climbing steps is not as important as the two on walking but has been included for consideration as in many cultures difficulty in this activity can have major impacts on an individual's ability to function. What ever the specific distances or number of steps chosen for inclusion, cognitive testing must be done to assure that the descriptors are understood in a similar way by all respondents.

45. To avoid respondent confusion, the question should include wording to indicate that the activities should be performed without the use of aids but cognitive testing should be conducted to determine if the inclusion of this wording has the desired effect or even if it is needed. Questions are not proposed about the use of assistive devices although such questions would be very useful and should be considered. Response categories should cover a range of difficulty levels, the exact wording should follow the template set for other domains.

Proposed items for measuring physical function – mobility:

By yourself, and without using any special equipment, how much difficulty do you have

...walking a few steps?

... walking a quarter of a mile/about 3 city blocks/xxx meters?

... walking up 10 steps without resting?

... running a short way (about 100 meters)

NO DIFFICULTY

ONLY A LITTLE DIFFICULTY

SOME DIFFICULTLY

A LOT OF DIFFICULTY

CAN'T DO AT ALL

Domain 2. Physical Functioning: DEXTERITY

46. Measures of dexterity are often, in the literature, included in a broader set of functional limitation measures. The first proposal for functional limitation measures was by Nagi (Nagi, 1976); he established the concept of Physical Performance, referring to sensory-motor functioning and described by limitation in activities such as walking, climbing, bending, reaching, hearing, etc. and built a scale for an epidemiological survey, individuals being classified as having none, some or great difficulty in the following:

- Difficulty standing for long periods
- Difficulty lifting or carrying weights of approximately ten pounds
- Difficulty going up and down stairs
- Difficulty walking
- Difficulty stooping, bending, or kneeling
- *Difficulty using hands and fingers*
- Difficulty reaching with either / or both arms.

47. The overview of the standardised dexterity measurement instruments and of the instruments used in the various National Health Interview Surveys resulted in the selection for analysis of the module recommended in the European Health Status Module (4 questions). The rationale behind the choice made for this Module is presented in the EUROREVES 2000 Report. An extract of the table presenting the related selected descriptors according to the criteria of selection is here presented:

Table. Selected descriptors according to criteria of selection

Descriptors	Number of covered categories (in the classification)	Simplicity and clarity of wording	Origins
Using fingers to grasp or handle a small object like a pen	<i>ONE: Agility</i> Shoulders, and Upper limbs (Grasping)	YES	US Longitudinal Survey of Aging (LSOA), 1984-90
Turning a tap	<i>ONE: Agility</i> Shoulders, and Upper limbs (Manipulating)	YES	British Disability Survey, 1996

NB: In the shape of a question, each descriptor uses a wording in term of capacity (“Can you”) and indicates the level of assistance (“by yourself and not using aids”).

Source: *EUROREVES 2000*

48. As shown in the table, and according to the review of the literature and through the analysis of the already existing international recommendations, the two main dimensions considered to measure dexterity (meant as agility using hands and fingers) are Grasping and Manipulating.

49. The criteria used by the EUROREVES group to select the best possible descriptor for each category were:

- Number of European Union countries using the question;
- Question following OECD or WHO-Europe recommendations;
- Number and type of categories of abilities covered by the question;
- Wording in terms of capacity rather than in terms of performance;
- Relationships with assistance (special aids or human assistance);
- Simplicity and clarity of wording.

European Health Status Module

1.Can you use your fingers to grasp or handle a small object like a pen without any difficulty and without any aids?

Yes

No

(We are not interested in subjects being able to use specially adapted implements but general, everyday ones. Disabled people may have special taps but the question relates to any tap not their own.)

If no:

Can you use your fingers to grasp or handle a small object like a pen with aids?

Yes

No

Has no aids

2.Can you turn on a tap or unscrew the lid of a jar of coffee without any difficulty and without any aids?

Yes

No

If no:

3.Can you turn on a tap or unscrew the lid of a jar of coffee with aids?

Yes

No

Has no aids

Comments:

The module allows distinguishing dexterity with and without the use of aids. It is possible in this way to measure capacity. However, if we are not necessarily interested in the measures that consider the use of aids, the module as it is could be redundant.

Proposed Items for Measuring Dexterity

50. The proposed instrument is selected from the European Health Status Module.

The duration of the recall period is not included as we are only interested in dexterity as a permanent aspect of the health state.

It could be useful to add, in each country, different examples to make clearer what is intended with the term aids as eventual comprehension problems maybe raised due to translation of this term in different languages.

Specific Preamble:

“The next questions ask about [your/his/her] ability to use [your/his/her] fingers and hands. When answering these questions, please **do not consider temporary problems** .

DEX-1. Can you use your fingers to grasp or handle a small object like a pen without any aids?

Yes, with no difficulties

Yes, with some difficulties

No

(We are not interested in subjects being able to use specially adapted implements but general, everyday ones. Disabled people may have special taps but the question relates to any tap not their own.)

If no:

DEX-2. Can you use your fingers to grasp or handle a small object like a pen with aids?

Yes

No

Has no aids

DEX-3. Can you turn on a tap or unscrew the lid of a jar of coffee without any difficulty and without any aids?

Yes, with no difficulties

Yes, with some difficulties

No

If no:

DEX-4. Can you turn on a tap or unscrew the lid of a jar of coffee with aids?

Yes

No

Has no aids

Domain 3. Vitality and Fatigue

51. Vitality/Fatigue was already proposed, since the first meeting in Geneva (May, 2004), as a domain of “functioning” that should be part of a set of core domains and then it was included in the paper “Identification of domains” prepared by Canada. In this paper it was already underlined that this domain may have structural dependence with the domains of affect and sleep. The concept of vitality/fatigue was also clarified in terms of tiredness, lack of energy or feeling rested and refreshed.

52. In the ICF Vitality is classified as Energy/Vitality in Body Functions under chapter 1 Global Mental Functions (b130) and more precisely (b1300) as energy level: mental functions that produce vigor and stamina.

53. The analysis of the surveys included in the HIS/HES database has showed the following main points:

- vitality/fatigue is not investigated in all the surveys/countries
- when investigated, from 1 up to 4 questions are used.
- when 4 questions are used they are generally part of the SF-36, with slight differences in the wording or answer categories among countries.
- time reference: almost all the surveys provide a time reference, the most commonly used is the “past 4 weeks/last 30 days/past month”.
- Answer categories: most of the surveys focus on the time dimension (“how much of the time...”), more than on the graduation in severity due to the functional limitation.

The overview of the standardised vitality/fatigue measurement instruments and of the instruments used in the various National Health Interview Surveys resulted in the selection for analysis of two main instruments: the module proposed for the European Health Status Module (4 questions) and the question used in the World Health Survey.

European Health Status Module

Questions: How much, during the past 4 weeks:

- Did you feel full of pep?
- Did you have a lot of energy?
- Did you feel worn out?
- Did you feel tired?

[Give showcard with the possible responses:]

Answer categories:

- All the time
- Most of the time
- Some of the time
- A little of the time
- None of the time

Comments:

The above questions are the SF-36 item on vitality. The SF-36 has been translated into many languages and the translation has been validated in the IQOLA project³².

³² The International Quality of Life Assessment (IQOLA) Project was established in 1991 to translate and validate the SF-36 Health Survey and other measures of patient-reported outcomes for international use. Initially, the IQOLA

The SF-36 questions on Vitality, together with questions on Social Functioning, Role- Emotional and Mental Health, are used to build a general measure of Mental Health.

The four questions have been considered as reference instrument also in the EUROHIS project (Nosikov A., Gudex C. (2003)).

This set of questions is included in the European Health Status Module that is right now under the testing process. It should be considered that the answer categories used in this module are 5, while in SF-36 they are 6: “a good bit of the time” is omitted in the European Health Status Module.

The concept of “vitality” (energy and fatigue level) is investigated by two positive (full of pep, a lot of energy) and two negative items (worn out, tired). The “alternation of positive and negative statements is a key part of the construction of an instrument that collects attitudinal data, as it mitigates against “response set”, i.e. it reduces the chance that the respondents will give the same answer to all questions. Essentially, this technique makes the respondent think about each question and the response categories.” (Meltzer, 2003).

In the questions there is not specific reference to the use of medicines; therefore we are not able to distinguish whether the answer is influenced by the use of medicines. (IS IT RELEVANT TO CONSIDER THE USE OF MEDICINES FOR THIS DOMAIN?).

World Health Survey

Question:

Q.2081 Overall in the last 30 days, how much of a problem did you have due to not feeling rested and refreshed during the day (i.e. feeling tired, not having energy)?

Answer categories: None
 Mild
 Moderate
 Severe
 Extreme

Comments:

This question assesses severity level of problems in performing due to not feeling rested or refreshed.

The question is focused on the impact produced by the lack of vitality/fatigue more than on the status of health itself.

The formulation of the question seems to be too complex as it requires to focus first on the presence of the condition of tiredness, and at the same time to evaluate the impact of this condition in everyday life.

In the question there is not specific reference to the use of medicines; therefore we are not able to distinguish whether the answer is influenced by the use of medicines. (IS IT RELEVANT TO CONSIDER THE USE OF MEDICINES FOR THIS DOMAIN?).

Proposed Items for Measuring Vitality/Fatigue

54. The proposed instrument is adapted from the European Health Status Module. This instrument seems to be the most simple to understand for the respondents. It is already available in several languages and it is being tested in several countries. The instrument seems to be the one that best allows to measure vitality and fatigue, as it is well balanced between negative and positive items (Meltzer, 2003).
55. Our proposal keeps only 2 of the 4 questions: the one on frequency of energy and the one on frequency of feeling tired. Moreover, the two questions on frequency are integrated with two VAS questions on intensity of energy and intensity of tiredness.
56. It is necessary to evaluate the opportunity to include in the preamble a reference to the use of medication, as suggested by Canada in the working document on conceptual and logistic issues in item construction. In fact, even if including reference to the use of medication would allow to better measure capacity, it would change the SF-36 questions not allowing to use the validation work done on it.
57. If it becomes necessary to cut the number of questions for this domain, it should be considered that, in the SF12, only one of the 4 questions is used to measure vitality, that is: **How much, during the past 4 weeks did you have a lot of energy?**

Specific Preamble:

“The next few questions ask about [your/his/her] vitality and fatigue.”

VIT-1 How much, during the past 4 weeks, did you have a lot of energy?

- All the time
- Most of the time
- Some of the time
- A little of the time
- None of the time

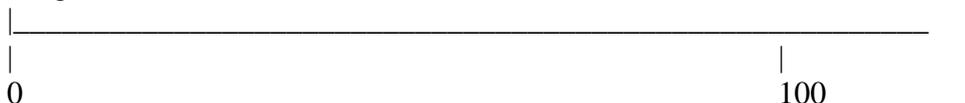
VIT-2 How much, during the past 4 weeks, did you feel tired?

- All the time
- Most of the time
- Some of the time
- A little of the time
- None of the time

VIT-3 Thinking about the past 4 weeks, on a scale from 1-100 how intense was the level of energy on average: 0 is lowest level and 100 is the highest level imaginable.

no energy at all
imaginable

highest level of energy



VIT-4 Thinking about the past 4 weeks, on a scale from 1-100 how intense was the level of tiredness on average: 0 is lowest level and 100 is the highest level imaginable.

no tiredness at all
imaginable

highest level of tiredness



0

100

Domain 4. Affect

58. The concept of affect is defined by the Task Force (TF) in terms of emotional state, with particular emphasis to the aspects of happiness and depression.
59. Affect is one of the domains of the Emotional area indicated in the draft paper on identification of domains as well as psychological functioning, anxiety, self esteem, relaxation and leisure.
60. The definitions about emotional functions given in ICF are used as a reference³³.
61. Analysing the instruments in use for measuring the emotional state, issues the overlapping of this domain with other domains also chosen for this work e.g. anxiety, vitality/energy and social relationships (see draft of identification of domains).
62. Regarding **affect** two main aspects are distinguished: positive feelings/psychical state and negative feelings/psychical state (Bradburn N., 1969). These are the two main components that are used for measuring affect.
63. Respecting the purposes of the TF to construct a parsimonious set of questions for each domain, in the process of selection of questions on happiness and depression, it was decided to exclude any reference to specific aspects.
64. For example, for depression, several other items included in measurement scales used in psychiatry are excluded from analysis: suicide (several batteries of questions were built regarding suicidal thoughts, intention to commit suicide and eventually attempted suicide); sleeping disturbances; lose/gain of weight; anxiety; being treated by specialists; the use of specific medications, etc.
65. However, it must be noted that an extremely synthetic set of question makes it very difficult to detect the presence and the level of severity of a certain emotional state.
66. It should also be mentioned that, not making any reference to the use of medicines, the measure of affect that we would obtain, would not be a measure of capacity. In fact, it can be assumed that, even in the case, for instance, of a serious form of depression, pharmaceutical treatments could modify the perception of the emotional state. In this case the instrument used would allow to measure the performance of the emotional state rather than the capacity.
67. The overview of the standardised affect measurement instruments and of the instruments used in the various National Health Interview Surveys resulted in the selection for analysis of three main instruments: the module proposed for the European Health Status Module, a selection of questions from the CIDI-short form (Composite International Diagnostic Interview) and the question used in the World Health Survey.

³³ B152: Emotional functions: specific mental functions related to the feeling and affective components of the process of the mind. Inclusions: functions of appropriate emotions, regulation and range of emotion, affect, sadness, happiness, love, fear, anger hate, tension, anxiety, joy, sorrow, ability of emotion; flattening of affect.

European Health Status Module

1. How much, during the past 4 weeks have you felt so down in the dumps, nothing could cheer you up?

[Give showcard with the possible responses:

All of the time

Most of the time

Some of the time

A little of the time

None of the time]

2. How much, during the past 4 weeks have you felt calm and peaceful?

[Give showcard with the possible responses:

All of the time

Most of the time

Some of the time

A little of the time

None of the time]

3. How much, during the past 4 weeks have you felt down-hearted and depressed?

[Give showcard with the possible responses:

All of the time

Most of the time

Some of the time

A little of the time

None of the time]

4. How much, during the past 4 weeks have you been happy?

[Give showcard with the possible responses:

All of the time

Most of the time

Some of the time

A little of the time

None of the time]

5. Would you describe yourself as being usually ... :

happy and interested in life

somewhat happy

somewhat unhappy

unhappy with little interest in life

so unhappy that life is not worthwhile

Comments:

The module is widely used in general population surveys in several countries. The first 4 questions are part of the module on Mental Health of SF-36, widely used and tested in known-groups of patients. The SF-36 has been translated into many languages and the translation has been validated in the IQOLA project³⁴.

Moreover, it is worth reminding that the European Health Status module is under testing.

³⁴ The International Quality of Life Assessment (IQOLA) Project was established in 1991 to translate and validate the SF-36 Health Survey and other measures of patient-reported outcomes for international use. www.iqola.org

The whole MH – mental health module from SF-36, includes also the question “How much, during the past 4 weeks, did you feel very nervous?”, which could however be included in the anxiety domain. Having the entire MH-mental health module of SF-36 could allow the use of the analysis techniques already available for the instrument.

Moreover the module, as the whole SF-36 questionnaire, has the advantage of being easy to administer and different data collection techniques can be used (it can be self-completed, it can be used in both telephone and face to face interviews).

The time reference “the last 4 weeks” could be to be too short, as answers may be influenced by recent events or temporary changes of the mood. However, as mentioned in the working document on conceptual and logistic issues in item construction prepared by Canada, it needs to be taken into account that a longer reference period could lead to the risk of focusing only on the most intense events.

Most importantly, several tests on SF-36 showed a robustness of the instrument in terms of empirical validity (groups of patients with known differences in their clinical and psychical conditions) (Cassileth et al., 1984; Dupuy, 1984; Smith, Monson, Ray, 1986), predictive evidence of validity (Ware et al. 1984), construct validity based on factor analysis (Cassileth et al., 1984; Ware, Davies-Avery, Brook, 1980) and correlation analysis with other health state measures (Cassileth et al.,1984; Dupuy, 1984; Nelson et al. 1983; Read et al. 1987).

Moreover, including the last question (Andrews item on happiness) (Andrews F. and Withey S., 1976) could help us understanding whether the referred emotional state is due to temporary problems or not.

CIDI (Composite International Diagnostic Interview) short form (partial)

1. During the past 12 months, was there ever a time when you felt sad, blue, or depressed for 2 weeks or more in a row?
Yes
No (*go to next section*)
Volunteered: I was on medication /antidepressants
2. For the next few questions, please think of the 2-week period during the past 12 months when these feelings were the worst. During that time, how long did these feelings usually last?
All day long
Most of the day
About half of the day
Less than half of a day
3. During those 2 weeks, did you feel this way...
Every day
Almost every day
Less often
4. During those 2 weeks did you lose interest in activities, work or hobbies that usually give you pleasure?
Yes
No

Comments:

Please note that, according to what said above on the necessity to reduce the number of questions for each domain, it was decided not to analyse the whole CIDI short-form but only a selection of it, that we believe could be used for our purposes. Questions on tiredness and loss of energy (part of the Vitality/fatigue domain), loss of weight, sleeping problems, difficulties in concentrating, loss of self-esteem, suicidal thoughts are not included. Moreover, in the original CIDI short-form, all questions are repeated for those who do not answer 'YES' to the first question but who refer to have lost interest in usual activities for 2 weeks over the last 12 months.

CIDI is a structured standardized diagnostic tool for trained non-clinical interviewers. It was designed for use in large-scale community epidemiological surveys and other settings without clinicians to make diagnoses. CIDI has also been designed to foster cross-cultural comparative research.

Domains of CIDI Short Form have recently been recommended for monitoring mental health in HIS settings in Europe (Lehtinen et al. 2001).

The CIDI measures are relatively new and until recently they have been rarely used in European countries.

The instrument seems to be of difficult comprehension for the respondents due to the double time reference (2 weeks in a row within the last 12 months).

It is mainly oriented to the measure of depressive episodes and it would confine the measure of affect only to negative feelings.

World Health Survey

Overall in the last 30 days, how much of a problem did you have with feeling sad, low or depressed?

- None
- Mild
- Moderate
- Severe
- Extreme

Comments:

This question assesses severity level of problems in performing due to feeling sad, low or depressed.

In the question there is not specific reference to the use of medicines; therefore we are not able to distinguish whether the answer is influenced by the use of medicines.

The question is focused on the impact produced by the condition of sadness or depression more than on the status of health itself.

The formulation of the question seems to be too complex as it requires to focus first on the presence of the condition, and at the same time to evaluate the impact of this condition in everyday life.

This question does not make any reference to the duration of the referred emotional state.

The time reference "the last 30 days" seems to be too short, as answers may be influenced by recent events or temporary changes of the mood.

It is mainly oriented to the measure of negative feelings.

Moreover, for standardizing the value given from each interviewee to the answer categories scale, the use of vignettes would be necessary, not being able to respect, in this case, the parsimony criteria.

Proposed Items for Measuring Affect

68. The proposed instrument is selected from the European Health Status Module. This instrument seems to be the one that best allows to measure emotional state in its wider meaning, not limiting the attention only to depressive state or negative feelings. It is also the most simple to understand for the respondents. Moreover, it is already available in several languages, it was widely tested and validated within the SF-36, and it is being tested in several countries.

69. However, our proposal keeps only 2 of the 4 questions: the one on frequency of happiness and the one on frequency of depression. Moreover, the two questions on frequency are integrated with two VAS questions on intensity of happiness and intensity of depression.

70. It is necessary to evaluate the opportunity to include in the preamble a reference to the use of medication, as suggested by Canada in the working document on conceptual and logistic issues in item construction. In fact, even if including reference to the use of medication would allow to better measure capacity, it would change the SF-36 questions not allowing to use the validation work done on it.

Specific Preamble:

“The next few questions ask about [*your/his/her*] emotional state.”

AFF-1. How much, during the past 4 weeks have you been happy?

[Give showcard with the possible responses:

All of the time

Most of the time

Some of the time

A little of the time

None of the time]

AFF-2. How much, during the past 4 weeks have you felt down-hearted and depressed?

[Give showcard with the possible responses:

All of the time

Most of the time

Some of the time

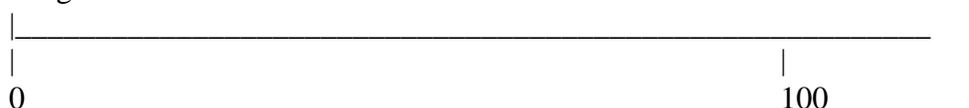
A little of the time

None of the time]

AFF-3. Thinking about the past 4 weeks, on a scale from 1-100 how intense was the level of happiness on average: 0 is lowest level and 100 is the highest level imaginable.

Not happy at all
imagine

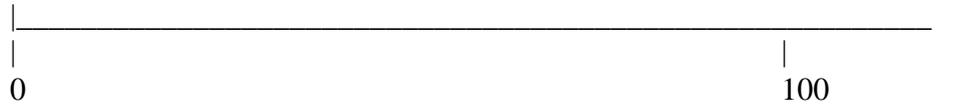
As happy as I could ever



AFF-4. Thinking about the past 4 weeks, on a scale from 1-100 how intense was the level of depression on average: 0 is lowest level and 100 is the highest level imaginable.

Not depressed at all
depression imaginable

The highest level of



0

100

Domain 5. Anxiety

71. In line with the criterion of item unidimensionality, it was determined that the Anxiety domain (i.e., worry, nervousness, fear) should be assessed completely separately from that of Affect (i.e., happiness – depression), rather than subsuming these concepts under a single item. A review of commonly used instruments, however, shows that anxiety is often assessed using double-barreled questions; in particular, it is often assumed to fall along the same continuum as depression. For example, as noted previously, the Anxiety and Depression domain used on the EQ-5D is of the following form:⁹

- 1. I am not anxious or depressed**
- 2. I am moderately anxious or depressed**
- 3. I am extremely anxious or depressed**

72. From an operational point of view, such a question might be confusing to respondents and they might ask interviewers for clarification (e.g., “If I am not depressed but am moderately anxious, which number should I choose: 1 or 2?”). Interpretation of data from such items might also be problematic, since it would be difficult to determine if a given participant was responding to the anxiety or depression component of the question. Although it is well known that depression and anxiety often occur in combination, they are distinguishable facets of mental health, both conceptually and empirically.^{10,11} For example, the widely-used Hospital Anxiety and Depression Scale (HADS) is designed to distinctly tap anxiety and depression, and factor analyses of the items have yielded two discrete yet correlated factors.¹²

73. Consistent with the view that Anxiety and Affect should be assessed separately on population-based surveys, the WHS employs the following two items:

Overall in the last 30 days, how much of a problem did [name of person] have with feeling sad, low, or depressed?

- 1 None**
- 2 Mild**
- 3 Moderate**
- 4 Severe**
- 5 Extreme**

In the last 30 days, how much of a problem did [name of person] have with worry or anxiety?

- 1 None**
- 2 Mild**
- 3 Moderate**
- 4 Severe**
- 5 Extreme**

74. Similarly, assessing Anxiety and Affect separately on the common instrument could provide further refinement in the assessment of mental health. Any correlation between Anxiety and Affect could be accounted for in later statistical analyses.

75. Presented here are three versions of the proposed questions for the Anxiety domain. The first, **Set A**, uses no recall period, that is, the phrase “Generally speaking...” was used to set the

question context. This was done to avoid confounding one's usual level of Anxiety with recent changes in environmental factors (e.g., loss of employment) or personal factors (e.g., chronic condition diagnosis), which might elevate it. Next, in light of potentially greater ease of translatability with using a specified recall period, a modified set of questions (**Set B**) is also included for consideration. The particular duration is left open at this point. **Sets A and B** include both frequency and intensity items, with the frequency items serving a filtering role. Finally, **Set C** attempts to implement the VAS approach described above, and also uses the recall period as in **Set B**.

76. All question sets measure three facets of Anxiety in order to help maximize measurement precision: worry, nervousness, and fear. The use of anxiety medication is also noted in the preamble for all question sets. In addition, questions across all sets are currently designed for self- as well as proxy-reporting, although it may be very difficult to give a proxy report on Anxiety. (The final instrument could drop the proxy component, if need be).

Proposed Items for Measuring Anxiety – Set A

Preamble: “The next few questions ask about [your/his/her] feelings of anxiety, that is, [your/his/her] feelings of worry, nervousness, or fear. When answering these questions, please take into account any medication you may be taking in order to relieve feelings of worry, nervousness, or fear.

ANX-1A. Generally speaking, how often do [you/he/she] have feelings of worry?
[INTERVIEWER: Read categories to respondent]

- 1 None of the time
- 2 A little of the time
- 3 Some of the time
- 4 Most of the time
- 5 All of the time

1, Don't Know, Refusal (Skip to ANX-3A)

ANX-2A. Generally speaking, how intense are [your/his/her] feelings of worry?
[INTERVIEWER: Read categories to respondent]

- 1 mild
- 2 moderate
- 3 severe
- 4 extreme

Don't Know, Refusal (Go to next Question)

ANX-3A. Generally speaking, how often do [you/he/she] have feelings of nervousness?
[INTERVIEWER: Read categories to respondent]

- 1 None of the time
- 2 A little of the time
- 3 Some of the time
- 4 Most of the time
- 5 All of the time

1, Don't Know, Refusal (Skip to ANX-5A)

ANX-4A. Generally speaking, how intense are [your/his/her] feelings of nervousness?
[INTERVIEWER: Read categories to respondent]

- 1 mild
- 2 moderate
- 3 severe
- 4 extreme

Don't Know, Refusal (Go to next question)

ANX-5A. Generally speaking, how often do [you/he/she] have feelings of fear?
[INTERVIEWER: Read categories to respondent]

- 1 None of the time
- 2 A little of the time
- 3 Some of the time
- 4 Most of the time
- 5 All of the time

1, Don't Know, Refusal (Skip to next section)

ANX-6A. Generally speaking, how intense are [your/his/her] feelings of fear?
[INTERVIEWER: Read categories to respondent]

- 1 mild
- 2 moderate
- 3 severe
- 4 extreme

Don't Know, Refusal (Go to next section)

Proposed Items for Measuring Anxiety – Set B

Preamble: “The next few questions ask about [your/his/her] feelings of anxiety, that is, [your/his/her] feelings of worry, nervousness, or fear. When answering these questions, please try and focus only on [your/his/her] feelings of anxiety during the past ___ days. Please also take into account any medication you may be taking in order to relieve feelings of worry, nervousness, or fear.

ANX-1B. Overall during the past ___ days, how often did [you/he/she] have feelings of worry?
[INTERVIEWER: Read categories to respondent]

- 1 None of the time
- 2 A little of the time
- 3 Some of the time
- 4 Most of the time
- 5 All of the time

1, Don't Know, Refusal (Skip to ANX-3B)

ANX-2B. Overall during the past ___ days, how intense were [your/his/her] feelings of worry?
[INTERVIEWER: Read categories to respondent]

- 1 mild
- 2 moderate
- 3 severe
- 4 extreme

Don't Know, Refusal (Go to next question)

ANX-3B. Overall during the past ___ days, how often did [you/he/she] have feelings of nervousness?

[INTERVIEWER: Read categories to respondent]

- 1 None of the time
- 2 A little of the time
- 3 Some of the time
- 4 Most of the time
- 5 All of the time

1, Don't Know, Refusal (Skip to ANX-5B)

ANX-4B. Overall during the past ___ days, how intense were [your/his/her] feelings of nervousness?

[INTERVIEWER: Read categories to respondent]

- 1 mild
- 2 moderate
- 3 severe
- 4 extreme

Don't Know, Refusal (Go to next question)

ANX-5B. Overall during the past ___ days, how often did [you/he/she] have feelings of fear?

[INTERVIEWER: Read categories to respondent]

- 1 None of the time
- 2 A little of the time
- 3 Some of the time
- 4 Most of the time
- 5 All of the time

1, Don't Know, Refusal (Skip to next section)

ANX-6B. Overall during the past ___ days, how intense were [your/his/her] feelings of fear?

[INTERVIEWER: Read categories to respondent]

- 1 mild
- 2 moderate
- 3 severe
- 4 extreme

Don't Know, Refusal (Go to next question)

Proposed Items for Measuring Anxiety – Set C

Preamble: “Now we are going to ask you to about [your/his/her] feelings of Anxiety during the past ___ days, that is, [your/his/her] feelings of worry, nervousness, or fear. For answering this set of questions, you will be asked to use a rating scale that goes from 0 to 100. 0 means that [you/he/she] were not worried, nervous, or fearful at all, and 100 means that [your/his/her] you were as worried, nervous, or fearful as you could possibly imagine. When answering these question, please try and focus only on [your/his/her] life during the past ___ days. Please also take into account any medication you may be taking in order to relieve feelings of worry, nervousness, or fear.

ANX-1C. Overall during the past ___ days, how would you describe [your/his/her] level of worry?

[INTERVIEWER: Present scale to respondent and say: “Please choose the number from 0 to 100 that best describes describe [your/his/her] level of worry during the past ___ days”]



Don't Know, Refusal (Go to next question)

ANX-2C. Overall during the past ___ days, how would you describe [your/his/her] level of nervousness?

[INTERVIEWER: Present scale to respondent and say: “Please choose the number from 0 to 100 that best describes describe [your/his/her] level of nervousness during the past ___ days”]



Don't Know, Refusal (Go to next question)

ANX-3C. Overall during the past ___ days, how would you describe [your/his/her] level of fear?

[INTERVIEWER: Present scale to respondent and say: “Please choose the number from 0 to 100 that best describes describe [your/his/her] level of fear during the past ___ days”]



Don't Know, Refusal (Go to next section)

Domain 6. Vision

77. Two are the main dimensions to be considered measuring the seeing capacity (EUROREVES 2000): seeing from far and seeing from close.

Using as a principal instrument the HIS/HES database, it can be identified that several health state measurement instruments use as tools the following concepts to measure the two dimensions:

Capacity seeing from far: (i) across the street (ii) four meters distance (iii) 20 meters distance

Capacity seeing from near: (i) arms length (ii) one meter distance (iii) newsprint letters.

Respecting the conceptual framework of the TF, the items taken into consideration and proposed are uni-dimensional (Paper 1 - Task Force).

Thus two separate questions should be used regarding the two different dimensions: ability seeing from far; ability seeing from close.

78. The prime problem faced regarding the domain Vision was deciding whether considering or not the use of glasses or other visual aids. Speaking about how to deal with the use of various aids, at the teleconference of 10 August 2005 it was discussed that glasses are considered part of the body whereas this is not the case for other kind of aids (e.g. for hearing aids). In fact, a review of several existing surveys and standardised health measurement instruments showed that the majority of the surveys measure the vision capacity using the formula “*with glasses or contact lenses, if necessary*” (see appendix).

79. Another feature here considered is the time reference. As discussed during the teleconference no specific agreement was made but there was the general consensus that if a time reference will be included it is preferred to include it in every question.

80. To obtain cross-cultural comparability it was decided to use well defined and objective descriptions regarding the measure of seeing difficulties. For instance: the reference to “across the street” was decided to be inappropriate as in different countries the average width of a street could be totally different. Thus we opted for the use of standard formulas as “4 meters away”, “one meter away” and “newsletter print”. In case the TF considers these formulations too difficult to be clearly perceived by respondents, it could be considered adding in brackets to the question specific examples, which could vary from country to country.

81. The overview of the standardised vision measurement instruments and of the instruments used in the various National Health Interview Surveys resulted in the selection for analysis of two main instruments, considered as the most representative and efficient in measuring vision. The two instruments are from the European Health Status Module (4 questions), and from the World Health Survey (3 questions). In the following section both instruments selected for review are presented, together with a proposal and the rationale for its selection.

European Health Status Module

1.Can you clearly see newspaper print without glasses, contact lenses or other visual aids?

Yes

No*

If no:

2.Can you clearly see newspaper print with your glasses, contact lenses or other visual aids?

Yes

No
Has no glasses or other aids

* if answer "I am blind or I cannot see at all", go to next section

3.Can you clearly see the face of someone 4 metres away (across a road) without glasses, contact lenses or other visual aids?

Yes

No

If no:

4.Can you clearly see the face of someone 4 metres away (across a road) with your glasses, contact lenses or other visual aids?

Yes

No

Has no glasses or other aids

(Visual aids includes magnifying glass, braille equipment)

Comments:

The above set of questions permits a distinguished double measuring of the seeing ability: with the use and without the use of visual aids. Thus it is available a net measurement of the capacity. Variations of the 2nd and the 4th questions are found in 14 surveys – see appendix (using the database HIS/HES).

However as we are not necessarily interested in the measure of vision without the use of visual aids, the 4 questions could be redundant. The reference to ‘across the road’ could mean different things in different countries.

World Health Survey

Do you wear glasses or contact lenses?

(If Respondent says YES to this question, preface the next 2 questions with "Please answer the following questions taking into account your glasses or contact lenses".)

Yes

No

In the last 30 days, how much difficulty did you have in seeing and recognizing a person you know across the road (i.e. from a distance of about 20 meters)?

None

Mild

Moderate

Severe

Extreme/Cannot do

In the last 30 days, how much difficulty did you have in seeing and recognizing an object at arm's length or in reading?

None

Mild

Moderate

Severe

Extreme/Cannot do

Comments:

Question number 1, so associated with the questions measuring the two main aspects of vision (from far and from close) allows to be able to measure the visual ability with or without the use of visual aids.

The reference distance of 20 meters seems to be too long and not in line with what is used in most national surveys (4m).

The measure across the road could mean different things in different countries.

The use of the term reading in question number 3 could be misleading in the case one person cannot read, we are only interested in the visual capacity to see small print characters. Moreover question 3 measures 2 separate concepts at a time, arm's length and reading.

Reference to the last 30 days does not seem to be appropriate as this could be misleading in the case of temporary problems.

Proposed Items for Measuring Vision

The proposed module combines the filter question on the use of glasses from the WHS (the use of other visual aids is added in the formulation of the question) with the two questions from the Health Status Module (considering the use of glasses). These two questions are modified slightly to provide a higher degree of measurement precision. In the question measuring visual ability from 4 metres the reference to the example 'across the road' is eliminated.

The duration of the recall period is not included as we are only interested in vision as a permanent aspect of the health state.

Specific Preamble:

"The next few questions ask about [your/his/her] visual ability; please do not consider temporary problems.

VIS-1. Do [you/he/she] wear glasses, contact lenses or other visual aids?

(If Respondent says YES to this question, preface the next 2 questions with "Please answer the following questions taking into account your glasses, contact lenses or other visual aids ".)

Yes

No

I am /he is /she is blind or I/he/she cannot see at all *(go to next section)*

VIS-2. How clearly can [you/he/she] see newspaper print?

Perfectly clearly

Very clearly

Somewhat clearly

Not very clearly

Not at all clearly

VIS-3. How clearly can [you/he/she] see the face of someone 4 metres away?

Perfectly clearly

Very clearly

Somewhat clearly

Not very clearly

Not at all clearly

(Visual aids includes magnifying glass, braille equipment)

Domain 7. Hearing

82. Hearing capacity can be evaluated on several different dimensions including the ability to hear sound at different levels of loudness or at different pitches and the degree to which hearing is affected by background noise. While it might be difficult for respondents to accurately describe the nature of their hearing loss, survey questions often try to tap into these various dimensions of hearing by describing the surroundings and the type of sound. While the intent is to standardize the stimuli, it is often difficult to do so as the descriptors used are themselves not standard and often do not describe all aspects of the environment. Obtaining information on the range of hearing problems would also require an extensive number of questions. When this is not possible, one or two questions are asked about specific situations.

Hearing is also closely associated with understanding speech and with communication. While some cognitive processes are involved, hearing limitations should be distinguished from other problems related to understanding language.

83. Hearing ability can be ascertained with or without the use of hearing aids. Hearing aids differ from glasses is that they are generally less successful in improving capacity overall and often while improving some aspects of hearing, do so at the detriment of other aspects. For example, hearing aids can increase volume but by doing so they make it harder to distinguish background noise from sources of sound. There is also more stigma associated with wearing a hearing aid in many cultures than with glasses and they are more difficult to obtain. It would be optimum to determine if a hearing aid is worn and to assess hearing with or without the aid. If this isn't possible, hearing is often assessed without the use of a hearing aid even if one is usually worn.

84. In selecting questions, it is important to decide what aspects of hearing to assess. Extensive question batteries are available that attempt to cover the range of hearing dimensions. One aspect of hearing that is often ascertained is the ability to hear normal conversation, that is conversation with one or a few persons. Sometimes it is stated that this conversation takes place in a 'quiet' room or that the speakers are face to face. A lack of difficulty in this kind of situation does imply that ability to hear in situations such as a classroom or train station.

Proposed Items for Measuring Hearing

One option is to use the questions from the European Health Status Module:

Can you distinctly hear what is said in a conversation with several people without a hearing aid or other aids for hearing? Yes / No

If no: Can you distinctly hear what is said in a conversation with several people with a hearing aid? Yes / No / Has no hearing aids

If yes: skip next question

Can you distinctly hear what is said in a conversation with one other person without a hearing aid? Yes / No

If no: Can you distinctly hear what is said in a conversation with one other person with a hearing aid? Yes / No / Has no hearing aids

Domain 8. Pain and Discomfort

85. "Pain is an unpleasant sensation that can range from mild, localized discomfort to agony. Pain has both physical and emotional components. The physical part of pain results from nerve stimulation. Pain may be contained to a discrete area, as in an injury, or it can be more diffuse, as in disorders like fibromyalgia. Pain is mediated by specific nerve fibers that carry the pain impulses to the brain where their conscious appreciation may be modified by many factors" (<http://www.medterms.com>)

86. From the ICF: Pain (b280-b289) - Sensation of unpleasant feeling indicating potential or actual damage to some body structure. Inclusion: sensations of generalized or localized pain, in one or more body part, pain in a dermatome, stabbing pain, burning pain, dull pain, aching pain; impairments such as myalgia, analgesia and hyperalgesia.

87. This section proposes items for measuring the domain of Pain and discomfort. . Scanning the web for suitable items and searching for candidates and making a review of established health state measurement instruments and questionnaire items from population health surveys (HIS/HES database: <https://www.iph.fgov.be/hishes/>) and the World Health Survey (WHS) (<https://www3.who.int/whs/P/instrumntandrel8293.html>), I acknowledge that pain is a difficult symptom to measure. It cannot be measured directly, but must be judged by the individual's response, which is subjective (influenced by sex, age, education etc). It is also a product of culture and condition. However, it is the subjective experience of pain that determines the consequences for the person and for society. And that is also the focus here.

Number of questions on pain

88. The aim should always be to try to cover all aspects of phenomena. Pain is often a symptom of a medical condition - an effect of cancer, muscular diseases, heart disease (ex. chest pain, angina), effect of injury etc. It can also be a symptom of stress (headache), migraine, or other types of more diffuse pain in muscles or skeleton. For women of a certain age pain can be associated with the monthly cycle of menstruation. Many surveys use questions for localized pain ie. ask if pain is situated in neck, shoulders, chest etc. However, with the limits that are set for number or proposed questions (2-4 questions) it is not possible to monitor the cause of pain, nor to monitor the whereabouts of pain. Experience shows that people that suffer from pain find it difficult to place the pain into the categories given (neck, arms, shoulders, head etc). A more generic approach is measuring the *intensity* of pain and *frequency* of pain. This will give an indication of the degree of suffering.

:) The restriction in number of questions necessitates a generic or global approach to pain.

Time reference

89. It is important to communicate clearly to the interview persons exactly what period of time is being studied: whether it is pain at this moment, pain last week, or pain the last 3 months/6 months. As the Task Force has not yet concluded on the inclusion of a time reference or recall period in the common instrument, and as this should be coordinated all though the survey, I will merely present some considerations here. However, choice of recall period is essential. The Canadian paper discusses the use of "usual" or "general" in survey questions - as *Are you usually free of pain or discomfort?* It is said that the "usual" version is employed in population health surveys, in order to avoid picking up the burdens of time-limited illnesses and conditions. A recall period for the survey questions would define the particular time span respondents should

use for estimating their level of functioning on each domain. Specific units of time are more comparable across cultures than for instance the concepts "usual" or "general". As referred to by Canada there should be a balance between "picking up" the, sometimes intense, experience of the past few days or the past week, and referring to a long period of for instance six months. The interview person's memory fades quickly when going back. It is important to ask ourselves what kind of information we want. What recall period is relevant for what information - the notion is that persons who struggle with pain over longer periods of time will have a problem with functioning and probably high correlation on depression. In Norway we have good experience with a three month recall on the symptoms questions in our HIS. However, we ask for *chronic or recurrent* problems.

:) Awaiting the workgroup to conclude on a more general basis as the number in recall periods through a questionnaire should be limited.

Uni-dimensionality

90. Methodologically it is said that a respondent should have *one* topic to relate to in each question, avoiding so called double-barreled questions. More than one topic is confusing and in measuring capacity IP might master one activity, but not the other. In the case of "pain and discomfort" it will shape-up the question and make it more distinct by not asking for discomfort. The discomfort-concept, that is fairly vague, blur the picture. In the McGills questionnaire discomfort(ing) is used as an answering category on a scale for describing the intensity of pain. From the McGill questionnaire (MPQ): (<http://www.proqolid.org/public/MPQ.html>):

"How intense has the pain been for you?"

no pain *mild* *discomforting* *distressing* *horrible* *excruciating*

:) Double-barreled questions may create ambiguity for survey respondents and can render it problematic to determine precisely what is being measured. Instead of asking about "pain and discomfort" we recommend one topic only: *pain*. I believe that not all will agree on this view and this should be an issue for further discussions in the TF.

Prosthetic (technical and medicinal)

91. There is agreement on measuring people's capacities and thus stay "within-the-skin". However, some domains may have to include information on prosthetics in order to discern what respondent are capable of doing without aids. Much pain can change behavior:

1. Seeking relief: what people do to relieve their pain (use of analgesics, hot water bottles, time spent resting each day etc
2. Functional assessment: to what extent does pain disturb function (eg. are you able to continue your work or other normal daily activities).

92. In some cases use of medication can give valuable information for getting an understanding of pain - its intensity and severity. Of course, the threshold for taking medication is very subjective and is related to how much pain one can stand, consciousness/beliefs, etc. It can also be problematic to differ between medicine on prescription and not on prescription. Country-specific rules and regulations will differ as to practice. The Canadian paper on selecting domains refers to a Canadian questionnaire where information about medication can be seen as to help gauge the severity of the feeling state of interest, such as the pain domain on the HU12 (<http://www.fhs.mcmaster.ca/hug/>):

1. *Free of pain and discomfort*
2. *Occasional pain. Discomfort relieved by non-prescription drugs or self-control activity without disruption of normal activities*
3. *Frequent pain. Discomfort relieved by oral medicines with occasional disruption or normal activities.*
4. *Frequent pain; frequent disruption of normal activities. Discomfort requires prescription narcotics for relief.*
5. *Severe pain. Pain not relieved by drugs and constantly disrupts normal activities.*

93. People with a chronic pain condition (ex. due to arthritis) will use pain killers on a regular basis and will probably have difficulty in answering unless it is specified in the questions that we ask for the severity of pain when not using medication. Canada suggests in their paper an approach to addressing the impact of medication by trying to standardize interpretation of the feeling state items across respondents ie. adding a statement to the preamble:

"When answering this question, please consider any medication you are using and the degree to which it relieved your Pain and Discomfort over the lastdays."

94. It is said that this should of course be pre-tested for comprehensibility and effectiveness.

:) The sensation of pain is difficult to measure, and instead of creating an ambiguity in the question it should state very clearly if the interview person shall answer taking into account medication or not. If there are available aids or medication that can better functionality or ease pain - is it then a health problem the TF want to capture? I await further discussions on this.

Item wording - response categories

95. The choice of rating scale is not straight forward. This issue will also depend on which strategy is chosen for the total survey questionnaire. However, I have included some information and thoughts presenting three types of rating scales:

1. Verbal rating scale (adjectives): These scales may be less reliable because adjectives do not necessarily mean the same to different individuals (nor across cultures). Some commonly used qualifiers are:

none slight moderate severe intense

or

none mild moderate severe extreme/cannot do (ref. ICF)

For defining the frequency of pain, ex.

all of the time most of the time some of the time a little of the time none of the time
(ref. NOS SF-36)

or

all the time (non-stop) daily (with pauses) several days per week, but not every day
 one or two times per week several times per month once a month less frequent

WHS vignettes: I will just comment shortly on the use of anchoring vignettes for cross-country comparability as this is elaborated more on in the Canadian paper. Norway did testing and data collection for the World Health Survey (WHS). Our experience with the use of vignettes (real life examples) was that it was difficult to administer in the interview and that the interview persons reacted negatively towards answering (especially regarding the most extreme examples). We did telephone interviews. The text was overwhelming both for the interviewers and the interview persons as one and each of the interview persons were presented with *a set of* real life examples. The recall period used in the vignettes were anticipated by the interview persons as artificial ("how can I know what ex. Malcom felt in the past 30 days?"). The rating of the vignettes on a scale from "no problem at all" to "extreme difficulty" used in examples where the outcome for the person was obvious ("How much difficulty does someone that is completely paralyzed have with moving around?") was asked to identify the extreme points on the scale. This seemed to irritate the respondents.

2. Numerical rating scale (numbers):

no pain worst pain imaginable
0 1 2 3 4 5 6 7 8 9 10

3. Visual Analogue Scale (VAS): Indicate pain levels on a straight line

no pain worst pain imaginable
 |-----|
 | |
 0 100

96. Both the numerical rating scale and the visual analogue scale are simple. It is said that the numerical rating scale is more easily used by people who are incapacitated by illness or have lower levels of education. A disadvantage is that it is an ordinal scale rather than a true value interval scale, thus there are no fixed relationship between the scores, even though boxes are spaced out at regular intervals. This means that the pain that scores 4 is not twice as severe as the one that scores 2. Statistically significant differences are likely not to be detected unless there are either large differences between groups, or large sample sizes. VAS is widely used and a proper ratio scale: this means that its two ends are rooted, and a doubling of the score does accurately reflect a doubling of the pain. Therefore t-test can be used in an analysis, so that significant differences can be identified with relatively small sample sizes or small differences between groups.

:) Visual analogue scales steer clear of misunderstandings and misinterpretations of specific concepts (ie. adjective scales). Adjective scales are difficult to translate and it is hard to ensure equal cut-off points between cultures.

Use of filters

97. Statistics Norway is concerned with dropping response rates in surveys. We are therefore very careful with sequences that might be anticipated by the respondents as repetition or questions of little relevance to them. The use of filters or skips in the questionnaire lowers the burden on each respondent. A sequence measuring for instance pain should therefore start out with a *screening question*: Do you have pain or have you had pain during period. yes/no. And then lead the yes-group through the more detailed questions.

2. I have mild physical discomfort or symptoms—for example, pain, ache, nausea, itching, etc.
3. I have marked physical discomfort or symptoms—for example, pain, ache, nausea, itching, etc.
4. I have severe physical discomfort or symptoms—for example, pain, ache, nausea, itching, etc.
5. I have unbearable physical discomfort or symptoms—for example, pain, ache, nausea, itching, etc.

PROPOSED ITEMS FOR MEASURING PAIN AND DISCOMFORT

PAIN *Have you had any pain during the past(days/weeks/month)s (all types of pain).*

- yes (if yes proceed to question `FREQ_PAIN`) ----->
 no

FREQ_PAIN *Thinking about the past(days/weeks/month)s: How often did you have pain?*

- all the time (non-stop)
 daily (with pauses)
 several days per week, but not every day
 one or two times per week
 several times per month
 once a month
 less frequent

Comment: Italy proposes to invert the scale of the answering categories on frequency of pain, starting from the less frequent and ending with the most severe. I cannot see how this is possible, or needed. On a more general basis I agree. The light/mild categories should be placed before the heavy/severe categories.

MED_PAIN *Have you used any medication to try to ease (the) pain during this period?*

- no
 daily
 now and again

INTENS_PAIN *Thinking about the last time you had pain: On a scale from 1-100 how intense was the pain: 0 is no pain and 100 is the worst pain imaginable.*

no pain worst pain imaginable
|-----|
| |
0 100

Comments: Italy has commented on the use of the proposed question on intensity of pain and find it problematic that the question is referring to "the last episode of pain". They argue that combinations of frequency and intensity should allow for evaluating the degree of pain. It would not be correct to extend the intensity referring to the last episode to all the episodes considered in the frequency question. The last episode might not be representative. My view is that a more general question on intensity of pain would be difficult to answer for the respondent and difficult to interpret. Alternative approaches would be to ask for intensity of the most prevalent or most frequent pain during a specific period (concurrent pain). Italy comments that even with this approach it will be difficult to combine frequency and intensity to classify different groups of respondents, as suggested by Canada in their comments. Yet another approach would be to eliminate the question.

Domain 9. Cognition

98. The Cognition domain was originally to be separated into the sub-domains of (a) memory and concentration and (b) problem-solving and thinking. However, a decision has been made to use 4 sub-domains to strive for more measurement precision. While a review of several existing surveys and standardized health state measurement instruments yielded a useful foundation of information for constructing Cognition-related items, it was found that many of these tools did not take a uni-dimensional approach to measurement. Rather, a common tendency among the instruments was to place different aspects of Cognition along the same continuum of functional capacity. For example, the Health Utilities Index Mark 3 (HUI3)¹³ simultaneously assesses memory, thinking, and problem solving ability – level 1 is **“Able to remember most things, think clearly and solve day to day problems.”** Similarly, the 15-dimensional Measure of Health Related Quality of Life (15D)³⁵ combines different aspects of Cognition within its Mental Function domain – level 1 is **“I am able to think clearly and logically, and my memory functions well.”** It should be noted that one of the instruments reviewed, the WHOQOL-100,³⁶ assesses four aspects of cognition (i.e., thinking, learning, memory, and concentration) separately (e.g., **“How well are you able to concentrate?”**).

99. In consideration of the above issues, the approach adopted here was to borrow useful concepts from existing tools to create items for measuring Cognition, and have each item measure its parent concept in as pure a fashion as possible. In particular, the uni-dimensional approach to assessing each aspect of cognition targeted here is in line with the format of the WHOQOL-100.³⁶ The aspects of Cognition being assessed, as well as item wording, are also borrowed from other tools such as the HUI3,¹³ 15D,³² and the WHS.⁶

100. Three versions of the proposed questions are presented here. The first, **Set A**, uses no recall period, that is, the word “usual” was used to set the question context. This was done to avoid confounding one’s usual abilities with the consequences of recent short-term illness onset (e.g., flu), or recent changes in environmental factors, which might impinge on cognitive functioning (e.g., increased noise pollution due to road or building construction around one’s home or workplace). However, in light of potentially greater ease of translatability with using a specified recall period, a modified set of questions (**Set B**) is also included for consideration. The particular duration is left open at this point. Further, for Sets A and B, the same response categories – “no difficulty, a little bit of difficulty, quite a bit of difficulty, a great deal of difficulty, and complete difficulty/unable to do it at all – are used, but with the recognition that other sets of qualifiers will likely be entertained in the process of trying to achieve a higher degree of international comparability. Finally, **Set C** attempts to implement the VAS approach mentioned above, and uses the recall period as in **Set B**. Questions in all sets are currently designed for self- as well as proxy-reporting, although it may be very difficult to give a proxy report on some of the concepts that the TF is working on, such as Affect and Pain and Discomfort. (The final instrument could drop the proxy component, if need be). Examples of potential filter questions are also provided.

³⁵ Sintonen H, Perkurinen M. A fifteen dimensional measure of health-related quality of life (15D) and its applications. In: Walker SR, Rosser RM, editors. *Quality of Life Assessment: Key Issues in the 1990s*. Dordrecht: Kluwer Academic Publishers; 1993. p. 185-195.

³⁶ The WHOQOL Group. The development of the WHO Quality of Life Assessment Instrument (The WHOQOL) In Orley J, Kuyken W, editors. *Quality of Life Assessment: International Perspectives*. Berlin: Springer-Verlag; 1994. p. 41-57.

Proposed Items for Measuring Cognition – Set A

Preamble: “The next few questions ask about [your/his/her] level of cognitive functioning, for example, [your/his/her] usual ability to concentrate, remember things, think, and solve daily problems.

COG-1A. How much difficulty [do/does] [you/he/she] usually have in concentrating on things?
[INTERVIEWER: Read categories to respondent]

- 1 no difficulty
- 2 a little bit of difficulty
- 3 quite a bit of difficulty
- 4 a great deal of difficulty
- 5 complete difficulty/unable to do it at all

Don't Know, Refusal (Go to next question)

COG-2A. How much difficulty [do/does] [you/he/she] usually have in remembering things?
[INTERVIEWER: Read categories to respondent]

- 1 no difficulty
- 2 a little bit of difficulty
- 3 quite a bit of difficulty
- 4 a great deal of difficulty
- 5 complete difficulty/unable to do it at all

Don't Know, Refusal (Go to next question)

COG-3A. How much difficulty [do/does] [you/he/she] usually have in thinking clearly and logically?

[INTERVIEWER: Read categories to respondent]

- 1 no difficulty
- 2 a little bit of difficulty
- 3 quite a bit of difficulty
- 4 a great deal of difficulty
- 5 complete difficulty/unable to do it at all

Don't Know, Refusal (Go to next question)

COG-4A. How much difficulty [do/does] [you/he/she] usually have in solving daily problems?
[INTERVIEWER: Read categories to respondent]

- 1 no difficulty
- 2 a little bit of difficulty
- 3 quite a bit of difficulty
- 4 a great deal of difficulty
- 5 complete difficulty/unable to do it at all

Don't Know, Refusal (Go to next question)

Proposed Items for Measuring Cognition – Set B

Preamble: “The next few questions ask about [your/his/her] level of cognitive functioning during the past ___ days, such as [your/his/her] ability to concentrate, remember things, think, and solve

daily problems. When answering these questions, please try and focus only on [your/his/her] life during the past ___ days.

COG-1B. Overall during the past ___ days, how much difficulty [did] [you/he/she] have in concentrating on things?

[INTERVIEWER: Read categories to respondent]

- 1 no difficulty
- 2 a little bit of difficulty
- 3 quite a bit of difficulty
- 4 a great deal of difficulty
- 5 complete difficulty/unable to do it at all

Don't Know, Refusal (Go to next question)

Relevant ICF Categories: b140 (Attention Functions); d160 (Focusing Attention)

COG-2B. Overall during the past ___ days, how much difficulty [did] [you/he/she] have in remembering things?

[INTERVIEWER: Read categories to respondent]

- 1 no difficulty
- 2 a little bit of difficulty
- 3 quite a bit of difficulty
- 4 a great deal of difficulty
- 5 complete difficulty/unable to do it at all

Don't Know, Refusal (Go to next question)

COG-3B. Overall during the past ___ days, how much difficulty [did] [you/he/she] have in thinking clearly and logically?

[INTERVIEWER: Read categories to respondent]

- 1 no difficulty
- 2 a little bit of difficulty
- 3 quite a bit of difficulty
- 4 a great deal of difficulty
- 5 complete difficulty/unable to do it at all

Don't Know, Refusal (Go to next question)

COG-4B. Overall during the past ___ days, how much difficulty [did] [you/he/she] have in solving daily problems?

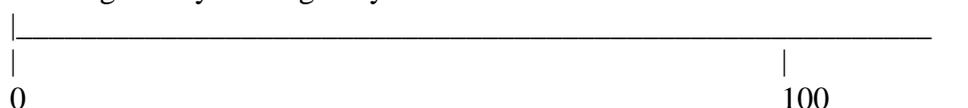
[INTERVIEWER: Read categories to respondent]

- 1 no difficulty
- 2 a little bit of difficulty
- 3 quite a bit of difficulty
- 4 a great deal of difficulty
- 5 complete difficulty/unable to do it at all

Don't Know, Refusal (Go to next question)

extreme/completely unable to think clearly and logically
thinking clearly and logically

no difficulty in



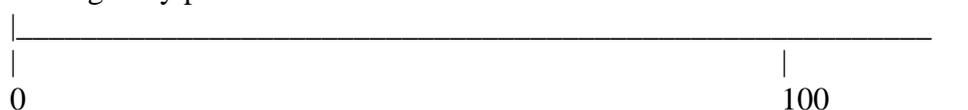
Don't Know, Refusal (Go to next question)

COG-4C. Overall during the past ___ days, how much difficulty [did] [you/he/she] have in solving daily problems?

[INTERVIEWER: Present scale to respondent and say: "Please choose the number from 0 to 100 that best describes how much difficulty [you/he/she] had in solving daily problems over the past ___ days"]

extreme/completely unable to solve daily problems
solving daily problems

no difficulty in



Don't Know, Refusal (Go to next question)

Potential Filter Questions For Use With Cognition Items

The following is one example of a filter strategy that might be considered for use with the suite of cognition items. Both the filter and follow-up questions are tailored to Set B above.

COG-1B-F. Overall during the past ___ days, did you have any difficulty [did] [you/he/she] have in concentrating on things?

[INTERVIEWER: Read categories to respondent]

1 no

2 yes

Yes (Go to question COG-1B)

No (Go to question COG-2B-F)

Don't Know, Refusal (Go to question COG-2B-F)

COG-1B. Overall during the past ___ days, how much difficulty [did] [you/he/she] have in concentrating on things?

[INTERVIEWER: Read categories to respondent]

1 a little bit of difficulty

2 quite a bit of difficulty

3 a great deal of difficulty

4 complete difficulty/unable to do it at all

Don't Know, Refusal (Go to next question)

Domain 10. Social Relationships

101. The approach taken here for assessing social relationships is quite distinct from that often used in other health state measuring instruments and population-based surveys. Specifically, the aim of the TF is to at producing items for measuring a person's *capacity* for establishing and maintaining social relationships, irrespective of the sheer number of opportunities for social interaction, whether he or she chooses to participate in social events, the actual extent of his or her social network, or his or her perceived level of social support. Perspectives have certainly varied regarding the usefulness of including social relationships in the assessment of health states. Although the comprehensive, classic definition of health put forward by the World Health Organization (WHO) includes social functioning,³⁷ many theorists and instrument developers have advocated a strictly "within-the-skin" approach to health state measurement.^{2,38} This perspective holds that the assessment of health states should be limited to physical and mental domains, with social functioning (e.g., role performance, engaging in activities with other people) considered as both a key health determinant and outcome. This perspective has merit, since explicitly incorporating social functioning into the assessment of health states could result in "double-counting" of mental and physical limitations that are socially expressed. For example, anxiety, depression, pain, and loss of mobility can all interfere with the execution of normal family and occupational functions. Thus, if all of the concepts just mentioned are combined when performing a quantitative assessment of a health state (e.g., computing a single global index or value representing the severity of the functional limitations involved - often a key objective in health state measurement), there may be considerable redundancy. On the other hand, the exclusion of social functioning from such assessments might not permit adequate measurement of health states in which the social limitations are themselves intrinsic to the individual (e.g., where fundamental difficulties understanding social cues and norms impinges on interactions with peers). Such basic limitations could work against a person's successful social integration and adaptation. Therefore, basic capacity concerning social interaction is an important aspect of human functioning and would seem to warrant direct measurement on population surveys.

102. It is important to note that the ICF²⁴ explicitly addresses individuals' basic abilities regarding social relationships. Specifically, Chapter 7 of the Activities and Participation component of the ICF is devoted to Interpersonal Interactions and Relationships, the categories of which are amenable to measurement in terms of either capacity or performance. For example, category **d720** (complex interpersonal interactions) encompasses the following concepts:

“maintaining and managing interactions with other people, in a contextually and socially appropriate manner, such as by regulating emotions and impulses, regulating verbal and physical aggression, acting independently in social interactions, and acting in accordance with social rules and conventions.”

103. These types of concepts offer a potentially rich foundation for instrument development. The ICF, however, is not a health state measurement tool in and of itself but rather offers a list of potential categories for assessment. However, a search for particular health state measuring

³⁷ World Health Organization. *Basic documents*. Geneva: World Health Organization; 1948.

³⁸ Ware JE, Brook RH, Davies AR, Lohr KN. Choosing measures of health status for individuals in general populations. *Am J Public Health* 1981; 71(6): 620-625.

instruments and survey items operationalizing the *capacity* for social relationships was not particularly fruitful. While instruments addressing social aspects of health are reasonably common, they primarily assess concepts such as social network size and social support levels. If social relationships are to be incorporated into the assessment of individuals' health states, a capacity- or ability-based approach seems essential for minimizing overlap with other domains, as well as for enhancing comparability across cultures (across which social role functions and expectations are often radically different).

104. A number of tools directly assess personal limitations in social relationships, but define these limitations as consequences of one's health state rather than intrinsic parts of it. For example, one of the questions posed on the SF-36 Social Functioning Scale is:²⁵

“How much of the time, during the past month, has your health limited your social activities (like visiting with friends or close relatives)?”

- 1 All of the time**
- 2 Most of the time**
- 3 A good bit of the time**
- 4 Some of the time**
- 5 A little of the time**
- 6 None of the time**

105. The following question on social relationships, employed on the various WHS modules,⁶ is somewhat more generic, treating social relationships as health state components rather than outcomes:

“Overall in the last 30 days, how much difficulty did you have with personal relationships or participation in the community?”

- None**
- Mild**
- Moderate**
- Severe**
- Extreme/Cannot do**

106. The current task was therefore to frame survey items to reflect peoples' functional ability concerning Social Relationships, and emphasize the social domain as being intrinsic to the individual. Only one general item was created, as it was felt that posing questions on different types of social relationships might diminish cross-cultural comparability by emphasizing environmental factors. For example, questions pertaining to interactions with different types of social contacts (e.g., immediate family members, relatives, friends, co-workers and strangers) might end up confounding true limitations in capacity with limitations on social interaction imposed by differing customs and/or laws. Of course, it would probably be unreasonable to expect that the influence of such factors on item responses could be eliminated entirely. However, it can likely be reduced by not specifying the types of social contacts involved. Further, this general approach potentially simplifies creating health state descriptions and conducting future preference elicitation exercises.

107. Three versions of the proposed questions are presented here. The first, **Set A**, uses no recall period, that is, the word “usual” was used to set the question context. This was done to avoid confounding one's usual abilities with the consequences of recent short-term illness onset

(e.g., flu), or recent changes in environmental factors which might impinge on social functioning (e.g., a recent move). However, in light of potentially greater ease of translatability with using a specified recall period, a modified set of questions (**Set B**) is also included for consideration. The particular duration is left open at this point. Further, for **Sets A and B**, the same response categories – “no difficulty, a little bit of difficulty, quite a bit of difficulty, a great deal of difficulty, and complete difficulty/unable to do it at all – are used, but with the recognition that other sets of qualifiers will likely be entertained in the process of trying to achieve a higher degree of international comparability. Finally, **Set C** attempts to implement the VAS approach mentioned above, and uses the recall period as in **Set B**. Questions in all sets are currently designed for self- as well as proxy-reporting, although it may be very difficult to give a proxy report on some of the concepts that the TF is working on, such as Affect and Pain and Discomfort. (The final instrument could drop the proxy component, if need be). Examples of potential filter questions are also provided. The relevant descriptive categories in the *International Classification of Functioning, Disability and Health*(ICF)²⁴ are also provided alongside the items, in order to show where the items fit within the general ICF framework.

Proposed Item for Measuring Social Relationships: Set A

Preamble: “Now I am going to ask you a couple of general questions about [your/his/her] social relationships. When answering these questions, I would like you to try and think about [your/his/her] social relationships with all of the people [you/he/she] meet[s] and talk[s] to in [your/his/her] daily life.”

SOCREL-1A. First, generally speaking, how much difficulty [do/does] [you/he/she] have in forming and maintaining social relationships?
[INTERVIEWER: Read categories to respondent]

- 1 no difficulty
 - 2 a little bit of difficulty
 - 3 quite a bit of difficulty
 - 4 a great deal of difficulty
 - 5 complete difficulty/unable to do it at all
- Don't Know, Refusal (Go to next question)

Proposed Item for Measuring Social Relationships: Set B

Preamble: “Now I am going to ask you a couple of general questions about [your/his/her] social relationships during the past ___ days. When answering these questions, please try and focus only on [your/his/her] life during the past ___ days.

SOCREL-1B. Overall during the past ___ days, how much difficulty [did] [you/he/she] have in forming and maintaining social relationships?
[INTERVIEWER: Read categories to respondent]

- 1 no difficulty
 - 2 a little bit of difficulty
 - 3 quite a bit of difficulty
 - 4 a great deal of difficulty
 - 5 complete difficulty/unable to do it at all
- Don't Know, Refusal (Go to next question)

Proposed Item for Measuring Social Relationships – Set C

Preamble: “Now I am going to ask you a couple of general questions about [your/his/her] social relationships during the past ___ days. For answering each of these questions, you will be asked to use a rating scale that goes from 0 to 100. 0 means that [you/he/she] had extreme or complete difficulty with social relationships, and 100 means that you had no difficulty with social relationships. When answering these questions, I would like you to try and think about [your/his/her] social relationships with all of the people [you/he/she] meet[s] and talk[s] to in [your/his/her] daily life.” Further, please try and focus only on [your/his/her] life during the past ___ days.

SOCREL-1C. Overall during the past ___ days, how much difficulty [did] [you/he/she] have in forming and maintaining social relationships?

[INTERVIEWER: Present scale to respondent and say: “Please choose the number from 0 to 100 that best describes how much difficulty [you/he/she] had with forming and maintaining social relationships over the past ___ days”]



Don't Know, Refusal (Go to next question)

Potential Filter Questions For Use With Social Relationships Items

The following is one example of a filter strategy that might be considered for use with the suite of social relationships items. Both the filter and follow-up questions are tailored to Set B above.

SOCREL-1B-F. Overall during the past ___ days, did [you/he/she] have any difficulty in forming and maintaining social relationships?

[INTERVIEWER: Read categories to respondent]

1 no

2 yes

Yes (Go to question SOCREL-1B)

No (Go to question SOCREL-2B-F)

Don't Know, Refusal (Go to question SOCREL-2B-F)

SOCREL-1B. Overall during the past ___ days, how much difficulty [did] [you/he/she] have in forming and maintaining social relationships?

[INTERVIEWER: Read categories to respondent]

1 a little bit of difficulty

2 quite a bit of difficulty

3 a great deal of difficulty

4 complete difficulty/unable to do it at all

Don't Know, Refusal (Go to next question)

108. *Communication*. Another aspect of social relationships that could be tapped is communication. An item attempting to capture the general capacity for making oneself understood in conversations is presented. In consideration of the usual trade-off between breadth of measurement and operational feasibility on population surveys, it was felt that such a general question would provide adequate coverage of communicative competency (i.e., be reflective of numerous underlying emotional, personality, cognitive, and behavioural factors without the need to explicitly tap each of these), as well as be a useful complement to the Social Relationships item. However, since communicative competency can be viewed as a partial mediator of the relationship between the Cognition and Social Relationships domains, there will probably be some degree of overlap among these domains.

109. There are versions with and without a reference period: A and B, respectively. The response categories remain the same (“no difficulty, a little bit of difficulty, quite a bit of difficulty, a great deal of difficulty, and complete difficulty/unable to do it at all”), but with the recognition that other sets of qualifiers will likely be entertained in the process of trying to achieve a higher degree of international comparability. Further, Set C implements the VAS approach. The questions are currently also designed for self- as well as proxy-reporting, although it may be very difficult to give a proxy report on some of the concepts involved (the final instrument could drop the proxy component, if need be). Potential filter questions are also provided. The relevant descriptive categories in the *International Classification of Functioning, Disability and Health* (ICF) ²⁴ are also provided alongside the items, in order to show where the items fit within the general ICF framework..

Proposed Item For Measuring Communication

COMM-1A. Second, how much difficulty [do/does] [you/he/she] usually have in making [yourself/himself/herself] understood in conversations?

[INTERVIEWER: Read categories to respondent]

- 1 no difficulty
- 2 a little bit of difficulty
- 3 quite a bit of difficulty
- 4 a great deal of difficulty
- 5 complete difficulty/unable to do it at all
- Don't Know, Refusal (Go to next question)

COMM-1B. Overall during the last __ days, how much difficulty [did] [you/he/she] have in making [yourself/himself/herself] understood in conversations?

[INTERVIEWER: Read categories to respondent]

- 1 no difficulty
- 2 a little bit of difficulty
- 3 quite a bit of difficulty
- 4 a great deal of difficulty
- 5 complete difficulty/unable to do it at all
- Don't Know, Refusal (Go to next question)

COMM-1C. Overall during the past ___ days, how much difficulty [did] [you/he/she] have in forming and maintaining social relationships?

[INTERVIEWER: Present scale to respondent and say: “Please choose the number from 0 to 100 that best describes how much difficulty [you/he/she] had with making [yourself/himself/herself] understood in conversations, over the past ___ days”]

extreme/completely unable to make
[myself/himself/herself] understood in conversations
[myself/himself/herself] understood in
conversations

no difficulty in making



Don't Know, Refusal (Go to next question)

Example of Possible Filter Question for Communication:

COMM-1B-F. Overall during the last ___ days, did [you/he/she] have any difficulty in making [yourself/himself/herself] understood in conversations?

[INTERVIEWER: Read categories to respondent]

1 no

2 yes

Yes (Go to question COMM-1B)

No (Go to question COMM-2B-F)

Don't Know, Refusal (Go to question COMM-2B-F)

COMM-1B. Overall during the last ___ days, did [you/he/she] have any difficulty in making [yourself/himself/herself] understood in conversations?

[INTERVIEWER: Read categories to respondent]

1 a little bit of difficulty

2 quite a bit of difficulty

3 a great deal of difficulty

4 complete difficulty/unable to do it at all

APPENDIX I for vision

MAIN CHARACTERISTICS	QUESTIONS	TOTAL NUMB. OF SURVEYS identical or similar questions
FOUR METERS (ACROSS THE ROAD) - without visual aids	Can you clearly see the face of someone <u>4 metres away</u> (across a road) without glasses, contact lenses or other visual aids? Visual aids includes magnifying glass, braille equipment. Yes No	6 (of which 3 use the question in combination with the following)
FOUR METERS (ACROSS THE ROAD) - with visual aids	Can you clearly see the face of someone 4 metres away (across a road) with your glasses, contact lenses or other visual aids? Visual aids includes magnifying glass, braille equipment. Yes No Has no glasses or other aids	17 (of which 3 use the question in combination with the previous)
20 METERS (across the road) - with visual aids	Please answer the following questions taking into account your glasses or contact lenses. In the last 30 days, how much difficulty did you have in seeing and recognizing a person you know across the road (i.e. from a distance of about 20 meters)? None Mild Moderate Severe Extreme/Can not do	2
One meter (at arm's length) - with visual aids: (Recommendations WHO-EURO)	Can you see well enough (with glasses or contact lenses, if necessary) to recognize a friend at a distance of one metre (at arm's length)? Yes No	9
One meter (at arm's length) - without visual aids (specified in a preamble)	Can you see well enough to recognise a friend at a distance of one metre (at arms length)? Yes No	5
One meter (at arm's length) + reading - with visual aids	Please answer the following questions taking into account your glasses or contact lenses. In the last 30 days, how much difficulty did you have in seeing and recognizing an object at arm's length or in reading ? None Mild Moderate Severe Extreme/Can not do	2
Newsprint - without visual aids	Can you clearly see <u>newspaper print</u> without glasses, contact lenses or other visual aids? Visual aids includes magnifying glass, braille equipment. Yes No	4 (of which 4 use the question in combination with the following)

MAIN CHARACTERISTICS	QUESTIONS	TOTAL NUMB. OF SURVEYS identical or similar questions
Newsprint - with visual aids	<p>Can you clearly see newspaper print with your glasses, contact lenses or other visual aids? Visual aids includes magnifying glass, braille equipment.</p> <p>Yes No Has no glasses or other aids</p>	18 (of which 4 use the question in combination with the previous)
Use of visual aids	<p>Do you wear glasses or contact lenses?</p> <p>No Yes</p>	9
One meter (at arm's length) + reading - with visual aids	<p>Please answer the following questions taking into account your glasses or contact lenses.</p> <p>In the last 30 days, how much difficulty did you have in seeing and recognizing an object at arm's length or in reading?</p> <p>None Mild Moderate Severe Extreme/Can not do</p>	2
Newsprint - without visual aids	<p>Can you clearly see <u>newspaper print</u> without glasses, contact lenses or other visual aids? Visual aids includes magnifying glass, braille equipment.</p> <p>Yes No</p>	4 (of which 4 use the question in combination with the following)

Source: HIS/HES database.

APPENDIX II: Related ICF Codes by Domain

For selected domains, a brief list of the relevant descriptive categories in the *International Classification of Functioning, Disability and Health* (ICF)²⁴ are provided in the table below, in order to show where the items fit within the general ICF framework. It is acknowledged that at this stage of the work there may not be an exact association between the ICF descriptive categories and the items presented here; and that the conceptual links with the ICF categories might not be perfectly preserved across different versions of the questions. However, the references to the ICF are intended to provide a common conceptual ground for further discussion and refinement of the items.

Table of Relevant ICF Categories for Selected Domains

Domain	Relevant ICF Categories
2: Dexterity	d440 (Using hands and fingers)
4: Affect	b152 (Emotional Functions)
5: Anxiety	b152 (Emotional Functions)
6: Vision	b210 (d110) (Seeing Functions)
8: Pain and Discomfort	b280-b289 (Pain)
9: Cognition	COG-1A/1B/1C - b140 (Attention Functions); d160 (Focusing Attention) COG-2A/2B/2C - b144 (Memory Functions) COG-3A/3B/3C - b1601 (Form of Thought); d163 (Thinking) COG-4A/4B/4C - b1646 (Problem-solving); d175 (Solving problems)
10: Social Relationships	d7200 (Forming relationships)
10: Communication	d330 (Speaking); d335 (Producing Non-verbal Messages) Could also cover d340 (Producing messages in formal sign language); d350 (Conversation)

References from the Italian contribution:

- European Health Interview and Health Examination Surveys Database. <https://www.iph.fgov.be/hishes/>. Accessed August 2005.
- Robine JM, Jagger C, Euro-REVES Group. Creating a coherent set of indicators to monitor health across Europe: the Euro-Reves 2 project. *European Journal of Public Health*, 2003.
- Nagi SZ (1976) An epidemiology of disability among adults in the United States. *Milbank Memorial Fund Q / Health Soc* 54:439-467.
- *EUROREVES 2000*, "Selection of a Coherent Set of Health Indicators - Final draft - A First Step Towards A User's Guide to Health Expectancies for the European Union" (Edited by J-M Robine, C. Jagger and V. Egidi).
- Task Force on Measurement of Health Status. Paper 1 (draft) – "Health as a multidimensional construct and cross – population comparability".
- Task Force on the Measurement of Health Status. Paper 2 (draft) – "Identification of domains".
- World Health Organization. *International classification of functioning, disability, and health*. Geneva: World Health Organization; 2001.
- The International Quality of Life Assessment (IQOLA) Project. www.iqola.org.
- Ware JE Jr, Snow KK, Kosinski M, Gandek B, SF-36 Health Survey Manual & Interpretation Guide.
- Nosikov A., Gudex C. (2003) "EUROHIS: Developing Common Instruments for Health Surveys", IOS press – WHO Regional Office for Europe.
- Meltzer H., (2003), "Development of a common instrument for mental health" in "EUROHIS: Developing Common Instruments for Health Surveys", IOS press – WHO Regional Office for Europe.
- Bradburn N., "The Structure of Psychological well-being"; 1969.
- Andrews, F. and Withey S., *Social Indicators of Wellbeing: American Perceptions of Quality of Life* (Plenum Press, New York), 1976.
- Cassileth BR, Lusk EJ, Strouse TB, Miller DS, Brown LL, Cross PA, Tenaglia AN, (1984) Psychological status in chronic illnesses: a comparative analysis of six diagnostic groups, *New England Journal on Medicine*, n.311, pp 506-511.
- Dupuy HJ, 1984, "The psychological general well-being (PGWB) Index", in Wenger NK, Mattson ME, Furberg CD, Elinson J (Eds), *Assessment of quality of life in clinical trials of cardiovascular therapies*, Le Jacq Publishing Company New York, pp. 170-183.
- Smith JR, Monson RA, Ray DC (1986), "Patients with multiple unexplained symptoms: their characteristics, functional health, and health care utilisation", *Archives of Internal Medicine*, n. 146, pp. 69-72.
- Ware JE Jr, Manning WG, Duan N, Wells KB, Newhouse JP (1984), "Health status and the use out patients mental health services", *American Psychologist*, n.39, pp. 1090-1100.
- Ware JE Jr, Davies-Avery A, Brook RH (1980), *Conceptualisation and measurement of health for adults in the Health Insurance Study. Volume VI: Analysis of relationship among health status measures*, The RAND Corporation, Santa Monica, CA.
- Nelson EC, Conger B, Douglas R, Gephart D, Kirk J, Page R, Clark A, Johnson K, Stone K, Wasson J, Zubkoff M (1983), "Functional health status level of primary care patients", *Journal of the American Medical Association*, n.249, pp.3331-3338.
- Read JL, Queen RJ, Hofer MA (1987), "Measuring overall health: an evaluation of three important approaches", *Journal of Chronic Disease*, n.40 (Suppl.1), pp. 7S-22S.
- Lehtinen V, Korkeila J, Morgan A, Salize H-J, Dalgard OS, Bijl R, Faria JS, Kovess V, (2001), Establishment of a set of mental health indicators for European Union (Final report): Commission of the European Communities and National Research and Development Centre for Welfare and Health STAKES.
- Task Force on Measurement of Health Status. "Summary of decisions and discussion from June 16, 2005 Teleconference"
- Task Force on Measurement of Health Status "Agreements reached during the 10 August 2005 Call of the Task Force on Health Status".
- Aromaa A, Koponen P, Tafforeau J, Vermeire C, Primatesta P, Marmot M, Kurth B, Gargiulo L, Lehto-Jarnstedt US, Raty S, *Health Surveys: evaluation and recommendations*, Helsinki 2003.