

***Introduction to the  
WHO Common Framework on Measuring and  
Reporting on the Health of Populations***

***Background Document for Session 4***

***UN-ECE and WHO Joint Meeting on Measuring Health Status  
hosted by Statistics Canada***

***Ottawa, 23-25 October 2000***



**Global Programme on Evidence for Health Policy  
World Health Organization  
Geneva**

# 1. Introduction and Background

The assessment of population health is an important activity in each country and at the global level. An array of data sources and methods are combined to estimate mortality, morbidity and other aspects of health. Collecting and rendering these statistics comparable across countries is one of the World Health Organization's longest standing mandates and has been recently characterized as an international public good [1]. Although constantly improving, rather well established data sources and estimation methods exist for mortality. This has not been the case for disability or other aspects of health. The improvement and standardization of methods to measure and compare health are therefore key components of the work of WHO, carried out in collaboration with Member States and research networks. This paper presents a common framework for the further development of this work, building upon current activities and addressing key challenges in the field of measuring and reporting on the health of populations.

## 1.1 State of measurement

This first section provides a brief overview of the current state of measurement of health, as opposed to disease or mortality.

### ▪ Standardized instruments to measure health

Ever since the 1947 WHO definition of health, people around the world have grappled with operational definitions and methodological approaches to measure health at the population level. The WHO definition clearly focused efforts on measuring the living state, rather than simply on the cause of death or mortality rates. Over the past 30 years, an array of standardized instruments have been developed to measure health, often through interview based surveys. Most efforts to assess different levels of health combined a range of questions drawn from four categories:

1. Items broadly reflecting the WHO definition (i.e., describing physical, mental and social aspects of health);
2. Items that are symptom-oriented or considered indicative of illness or morbidity;
3. Items that focus on the capacity or performance of functions, activities or roles as indirect ways to measure health status; and
4. Items that are concerned with adaptation to and coping with health conditions, including disabilities.

The first wave of standardized instruments assessed the most severe states of health, particularly among older age groups and individuals living within long-term care institutions. These measurement approaches focused on performance in different areas, such as questions on "Activities of Daily Living" (ADLs -- e.g. eating, getting in/out of bed, getting around inside, dressing, bathing or using the toilet) [2]. The second wave of instruments was developed with clinical and general populations in mind, and combined self-assessment of descriptions on different dimensions of health and of performance in different activities and roles. Data collected by these instruments are often presented as multiple dimensional profiles of health.

### ▪ Multi-dimensional approach to measure health

The distinct domains of health measured by thirteen of the most popular standardized instruments are noted in Table 1. The range of domains included across instruments reflect different practical ways to define and assess health. Although each instrument measures multiple domains of health, few use the same labels for domains or scales, or cover the same content in terms of the questions that make up each domain.

Not shown in Table 1, is that some domains include items focusing on a specific function, such as *vision*, while others encompass items assessing a range of complex functions and activities, such as *understanding and interacting*. Others may create a domain to focus on one aspect, such as performance of activities related to the home, while others may incorporate performance of activities or

roles at work, home or recreation into a single domain. The same applies to other domains, such as *eating* or *self-care*, with *self-care* usually incorporating questions addressing eating, bathing, dressing and other similar items.

Given these differences, Table 1 approximately represents the domains of the 13 instruments included. The main point is that no one standard set of domains and questions to describe health is used to facilitate comparison of multi-dimensional profiles of health status across populations. Nevertheless, most instruments include questions addressing general health, physical functioning and mental health state descriptions, as well as the performance of daily activities and the fulfillment of social or other normative roles [3].

**Table 1. Domains of 13 Generic Health Status Instruments**

Health Domains (multi-dimensional profile)	QWB '70	McM '76	SIP '76	QLI '81	NHP '81	FSQ '86	CP '87	Duke '90	SF-36 '92	HUI-III '95	WHO QOL '96	EQ 6D '99	WHO DAS II '99
<i>Overall Well-Being</i>													
<i>General Health</i>													
<i>Change in Health</i>													
<i>Physical Health</i>													
Activities/roles													
Work													
Home													
Recreation													
Ambulation													
Eating													
Energy/vitality													
Dexterity													
Hearing													
Mobility/fitness													
Pain/discomfort													
Self Care													
Sleep/Rest													
Speech													
Vision													
<i>Social Health</i>													
Activities/roles													
Communication													
Interaction													
Support													
<i>Mental Health</i>													
Activities/roles													
Alertness													
Anxiety/Depression													
Cognition													
Emotional status													
Outlook													
Self-esteem													
Understand/Interact													
<i>Handicap/Participation</i>													
<i>Environmental Context</i>													

QWB:Quality of Well-Being Scale, McM: McMaster Health Index, SIP: Sickness Impact Profile, QLI: Quality of Life Index, NHP: Nottingham Health Profile, FSQ: Functional Status Questionnaire, CP: COOP Charts for Primary Care Practice, Duke: Duke Health Profile; SF-36:Short-Form 36 Health Survey; HUI-III: Health Utilities Index Mark III; EQ6D: EuroQol 6 Domain Quality of Life Scale (5D excludes cognition), WHOQOL: WHO Quality of Life Bref Field Trial Version, WHODAS II: WHO Disability Assessment Schedule.

▪ **National Interview Surveys on health**

Over the past fifty years, a growing number of countries in industrialized and less industrialized countries are conducting nationally representative surveys dedicated to assessing the health of the population. Other national surveys include health topics and are often implemented on a

regular basis within household surveys. Examples include the National Health Interview Survey (NHIS) carried out annually in the United States since 1957 and the National Sample Survey (NSS) implemented routinely in India since 1950 that has incorporated specialized modules addressing health topics, such as those within Round 28 (1973-74), Round 35 (1980-81), Round 42 (1986-87) and Round 52 (1995-96).

One of the main advantages of data collected through household surveys is that they provide person- or household-based health statistics rather than data collected through health services or disease registries, which are episode- or event-based [4]. It is well established that episode- or event-based data are not representative of the population nor of conditions that are less well defined or do not have effective treatments. This is primarily due to heterogeneity in the access to health services and in health-seeking behavior within a population, as well as the uneven quality of routinely collected data within the health services [5].

#### ▪ **National Interview and Examination Surveys on health**

Household interview surveys that also include an examination component provide a more comprehensive profile of health status. Not surprisingly, these require substantially greater resources and tend to focus on clinical rather than functional measures of health status. A relatively few number of countries have conducted at least one nationally representative survey in the past decade combining household interviews with clinical examinations or laboratory assessments: these include Germany, Finland, Indonesia, Malaysia, Pakistan, United Kingdom and the United States. Older surveys exist, such as in Papua New Guinea (1963-64), Columbia (1964-65) and Egypt (1981-84) [6].

#### ▪ **Differences in survey topics and questions**

Despite increasing awareness of the need to standardize the collection and analysis of data, few standardized modules for national health surveys exist and are utilized across countries. This is so despite a range of research networks and international agencies that have initiated efforts to improve the relevant technical and analytical skills required to implement either household interview or interview and examination surveys.

Problems of comparability also exist within countries, between different geographic areas or time periods. How health is defined and measured varies widely. For example, in some countries the distribution and severity levels of health are equated with the types of risk factors or life style practices, the utilization of or level of satisfaction with health services and drugs, or even the level of health care expenditures estimated through surveys. Although interesting for other purposes, information on lifestyles, satisfaction or expenditures are not substitutes for information on health status. Furthermore, existing standardized modules developed by international agencies tend to focus on morbidity or disease classification rather than on describing health status. The result is that data focusing on morbidity serve primarily as inputs to “policies on curative care” [7]. Other surveys with standardized modules, such as the Demographic and Health Surveys conducted in some 65 countries have in the past interviewed only women.

Several detailed comparisons of the content and questions of national surveys addressing health have been recently published. Collectively, these show that questions specifically addressing health typically assess the severity and duration of functional or cognitive disability and the impact on usual activities. For example, in a review of 52 surveys conducted in the European Union, of the ten most common questions included in the 52 surveys, only questions on height and weight were considered comparable [8]. In a recent review of 30 surveys in 23 OECD countries, including non-EU countries as Australia, Canada, the Czech Republic, Korea, Japan, New Zealand, Switzerland and the United States, the main finding was that current differences in questions limit comparability only to those countries that have also incorporated standardized health status instruments in national health surveys, such as the SF-36 or EuroQol-5D [9].

WHO has recently reviewed an additional 64 surveys conducted within 46 developed and developing countries, and evaluated the types of questions focusing on describing health status rather than morbidity or disease classification [10]. In operational terms, this translated to questions covering the general state of physical or mental health, physical or cognitive functioning, other disability, as

well as the degree and duration of activity or role limitations. Table 2 lists example questions from five of the 64 surveys reviewed. The main finding is that across surveys considerable variation exists concerning the number and types of questions addressing the description of health.

**Table 2. Questions describing health, extracted from different national surveys**

<i>Survey Type</i>	<i>Questions</i>
<b>Integrated Household Survey – LSMS survey: Paraguay 1996</b>	<ol style="list-style-type: none"> <li>1. Have you been sick during the past 4 weeks?</li> <li>2. Have you been injured in the past 4 weeks?</li> <li>3. For how many days during the past 4 weeks were you unable to carry out your usual activities because of this illness or injury?</li> </ol>
<b>Longitudinal Integrated Household Survey – LSMS survey: Jamaica 1996, round 10</b>	<ol style="list-style-type: none"> <li>1. Have you suffered from an injury during the past 4 weeks?</li> <li>2. Have you suffered from an illness other than an injury in the past 4 weeks?</li> <li>3. Did this injury/illness begin in the past 4 weeks or exist before?</li> <li>4. For how many days during the past 4 weeks did you suffer from this illness of injury in the past 4 weeks?</li> <li>5. For how many days during the past 4 weeks were you unable to carry out your usual activities because of this illness or injury?</li> <li>6. Is a mental or physical disability preventing you from securing and maintaining employment?</li> </ol>
<b>Longitudinal Integrated household survey: Indonesian Family Life Survey 93-94, wave 1</b>	<ol style="list-style-type: none"> <li>1. In general, how is your health at this time?</li> <li>2. If you have to carry of heavy load for 20 meters, could you?</li> <li>3. If you have to sweep the house floor or yard, could you?</li> <li>4. If you have to walk for 5 kilometers, could you?</li> <li>5. If you have to draw a pail of water from a well, could you?</li> <li>6. If you have to bow, squat, kneel, could you?</li> <li>7. If you have to dress without help, could you?</li> <li>8. If you have to stand up from sitting position in a chair without help, could you?</li> <li>9. If you have to go to the bathroom without help, could you?</li> <li>10. If you have to stand up from sitting on the floor without help, could you?</li> <li>11. Have you ever experienced insomnia during the past 4 weeks?</li> <li>12. Have you ever experienced fatigue or exhaustion during the past 4 weeks?</li> <li>13. Have you ever felt short-tempered or hyper-sensitive during the past 4 weeks?</li> <li>14. Have you ever experienced bodily pains in the past 4 weeks?</li> <li>15. Have you ever experienced sadness during the past 4 weeks?</li> <li>16. Have you ever experienced anxiety or fear during the past 4 weeks?</li> </ol>
<b>Longitudinal Integrated household survey: European Commission Household Panel 1995, wave 2</b>	<ol style="list-style-type: none"> <li>1. How is your health in general?</li> <li>2. Do you have any chronic physical or mental health problem, illness or disability?</li> <li>3. Are you hampered in your daily activities by this physical or mental health problem, illness or disability?</li> <li>4. In the past 2 weeks: Have you had to cut down on any of the things you usually do about the house, at work or in your free time because of illness or injury?</li> <li>5. In the past 2 weeks: Have you had to cut down on any of the things you usually do about the house, at work or in your free time because of an emotional or mental health problem?</li> </ol>
<b>National health survey: Pakistan 1990-1994, Adult module</b>	<ol style="list-style-type: none"> <li>1. Would you say your health in general is excellent, very good, good, fair or poor?</li> <li>2. Have you passed/vomited worms during the last 3 months?</li> <li>3. Have you had a cough with phlegm in the past 14 days?</li> <li>4. In the past 14 days have you had loose stools?</li> <li>5. Do you experience sudden shortness of breath at night?</li> <li>6. Do you have pain in chest on exertion?</li> <li>7. Coming into a house at night would you have trouble seeing anything in the house?</li> <li>8. Do you have difficult hearing?</li> <li>9. If you had an accident (poisoning, fall, injury, burn etc.) in past 12 months, are you fully recovered, under recovery or recovered with some handicap?</li> </ol>

## 1.2 Major advances

It is worthwhile to highlight that the field of measuring and reporting on the health of populations has made considerable progress. These include conceptual advances, operational advances, as well as advances concerning the use of data collected. Key aspects of these advances are briefly touched on below.

- **Recognition that health is a multi-dimensional concept**

A health state is what an individual experiences and is a multi-dimensional concept, building on the WHO definition of health. Common domains used to describe health include physical functioning, cognitive functioning, pain and affect, among others. Every health state can be described on each domain of health. Descriptions of health states using domains provide a profile of health. A profile describing health thus differs from etiology-, symptom-, morbidity-, or disease-based classifications, such as entries within the International Statistical Classification of Diseases and Related Health Problems (ICD-10) [11].

For example, an individual with the disease label “acute paralytic poliomyelitis, wild virus, indigenous” would be classified within one entry of the ICD-10 – *viral infections of the central nervous system, acute poliomyelitis*. Once the disease is no longer present, any lasting effects of the disease that the individual experienced or sequelae of poliomyelitis, would be classified within another entry – under *sequelae of infections and parasitic diseases*. Whereas the same individual regardless of his or her disease status would have the level of his or her health assessed in each domain included within a multi-dimensional profile of health – such as within those domains listed in Table 1.

- **Recognition that validity and reliability must be estimated for each approach to measure health and within each population**

Standardized instruments to measure health on different domains have drawn upon a variety of methods to provide evidence that instruments meet criteria for validity and reliability, and that data may be meaningfully interpreted within a given population. Such estimates are required regardless of the survey mode, i.e., household interview, postal survey or internet survey, among other data collection modes.

Reliability refers to the stability of a measurement -- the extent to which an instrument or survey module will provide the same result on separate occasions. Validity is the extent to which an instrument measures what it is intended to measure, or more broadly, the range of interpretations that can be reasonably attributed to a measure. Overall, validity estimates for health status measures are reported less frequently than reliability estimates. Given that no gold standard measurement technique exists to measure health, an instrument is never “valid” based on a single test or automatically valid for other populations than the one it was originally developed and tested.

Rather, several approaches to estimate validity are usually cited as no single aspect of validity provides a definitive evaluation of an instrument [12]. Approaches to estimate validity also are used to gauge the interpretation of data collected. For example, criterion-based interpretation provides information on the differences between *known groups* within a particular context, such as individuals with mental illness and those with acute physical impairments. It is expected that these two groups will differ concerning their average score on a scale assessing mental health. These differences may provide guidelines on the meaning of scores measured on ordinal or interval scales. For example, the difference between the average scores observed in these two groups establishes a basis to interpret the size of large differences in scores [13].

- **Recognition that information on health is as important as information on mortality: Summary Measures of Population Health**

Summary measures of population health (SMPH) are measures that combine information on health and mortality, to represent the health status of a population as a single number [14]. Data obtained through standardized instruments or survey modules measuring health may provide information on the prevalence of different severity levels, and thus serve as an input to the calculation of SMPH.

For example, Disability-free Life Expectancy (DFLE), one type of SMPH, has been calculated for a number of countries since the 1980s using data collected from nationally representative surveys. DFLE incorporates a dichotomous weighting scheme, i.e., one that does not account for varying levels of severity. Based on a more sensitive weighting scheme that takes into account differences in the level of severity, the Disability-adjusted Life Expectancy (DALE, or healthy life expectancy) is another

SMPH and was estimated for publication within the recent *World Health Report* [15]. DALE includes data on health status from two sources [16]: (1) the prevalence of disabling sequelae for a wide range of diseases, morbidities and conditions obtained through a meta-analysis of epidemiological studies and an evaluation of the severity of these disabling sequelae; and (2) the severity level and distribution of physical and cognitive impairments and activity restrictions irrespective of morbidity or disease classification, obtained from nationally representative household surveys.

### **1.3 Key challenges**

Despite these advances, there remains several key challenges particularly related to the comparability of data on the severity and distribution of health across countries. Four of these challenges are briefly discussed below.

- **Need to agree on a common language to identify domains**

How to describe health in a multi-dimensional way, that is comprehensive and comparable, is a fundamental challenge. Differences in the domains used to describe health within each instrument or survey are large. It is therefore clear that no common language or classification system of health has been agreed upon or is being used as the basis for describing, measuring or reporting health (e.g., akin to the ICD for morbidity and cause of death). Terminology, including the definition of domains, also varies across instruments, surveys and applications.

- **Need to develop and test a common instrument**

A common instrument and standardized survey methods tested within nationally representative samples, in both developed and developing countries, would benefit all working on the measurement and comparison of health status across countries.

Differences across standardized health status instruments are as great as differences across household surveys. Not shown in Table 1 is that all instruments use different combinations of questions for each domain included, and these vary in range and depth, recall periods, type of response scales and scoring methods. Each of these differences may affect the distribution of responses, the comparison across among measures and the ability to interpret data. For example, within the SF-36 Health Survey, the physical functioning dimension includes ten items assessing whether the person has limitations engaging in vigorous activities such as running, lifting heavy objects, or participating in strenuous sports on the high end of the scale, to limitations in bathing or dressing, on the low end of the scale. Within the EuroQol Quality of Life Scale, each dimension is made up of a single question.

Currently, no single instrument or survey module is used across countries to assess health. There is no indication that any one of the available instruments will become the de facto standard across all regions in the near future. This is so despite selected, concentrated efforts. For example, the validity and reliability of the Short-Form 36 Health Survey has been estimated in some 10 developed countries through nationally representative surveys [17], and in a much larger number of countries within clinical and other sample groups. Another example includes the EUROHIS project [18]. This collaborative project within Europe has as its main goal to increase the comparability of survey modules addressing different domains of health within European health interview surveys: national surveys could incorporate these modules or adjust existing ones to improve the comparability of questions.

- **Need to facilitate integration of data on health within summary measures of population health**

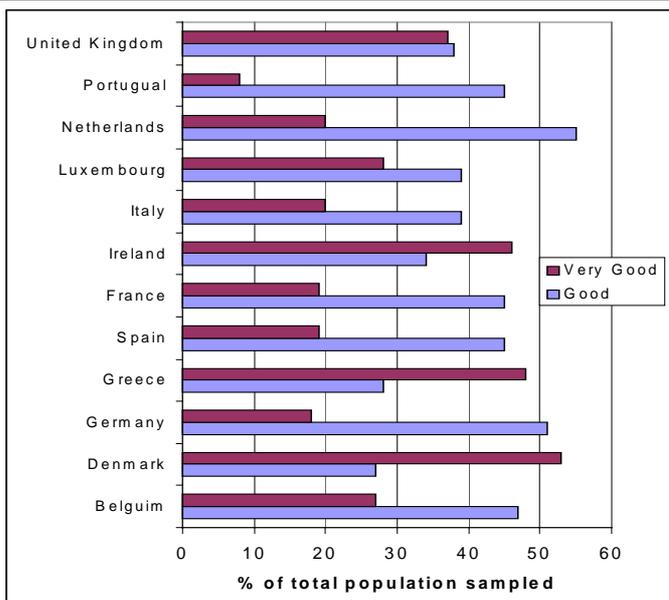
It is becoming widely accepted that mortality alone does not provide a comprehensive picture of a population's health status; in parallel, a growing body of data on severity and distribution of health from nationally representative surveys exist. Yet there remains a gap between the collection of data on health and its input within summary measures of population health (SMPH).

Some efforts to improve the link between data collection and integration with SMPH exist. For example, the Network on Health Expectancy (Réseau Espérance de Vie en Santé or REVES) has recently reviewed survey data sources for the calculation of Disability-free Life Expectancy (DFLE) and recommended criteria towards improving survey questions in five areas or domains of health: physical and sensory functional limitations, activity restrictions in daily life, limitations in usual activities, general health status, and mental health indicators [19]. Although this is a useful and important step, the methods focus on defining a single threshold of disability or ill-health, and do not take into account the range of severity levels of health assessed within population representative surveys.

- **Need to identify and improve cross-population comparability**

Cross-population comparability refers to the property that a similar response on an item or domain score maps to the same true level of health in different populations. In addition to the classical psychometric properties of instruments (e.g. reliability and validity) and other issues concerning survey implementation (e.g., sampling frame, missing data), the issue of cross-population validity emerges as an essential requirement for international comparisons. The instruments or survey modules mentioned above may be reliable and valid for use within a given country or population. However, even if the same instrument and questions are used among different populations, the data may still not be comparable.

Figure 1. Proportion of population  $\geq 16$  years of age, reporting very good and good general health, 12 European countries, Eurostat 1997 (European standard population age-standardization)



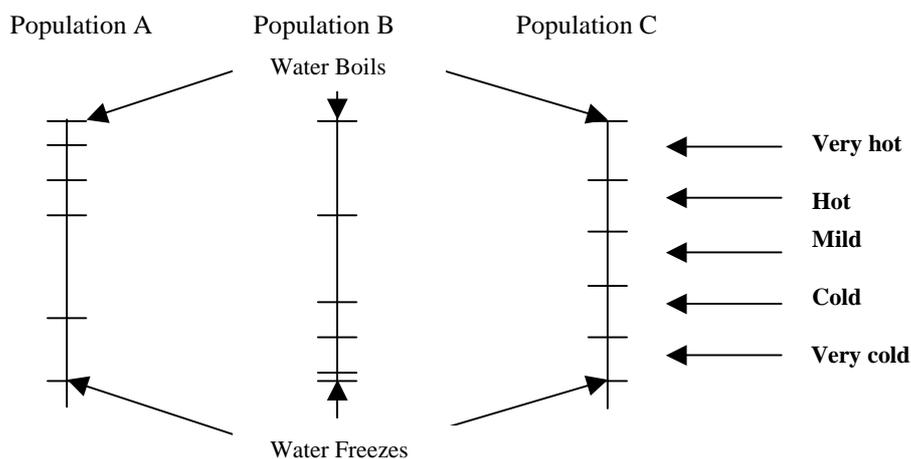
For example, results reported from the European Community Household Panel survey, relying on self-reported data collected in 12 countries using identical methodologies, are illustrative. Figure 1 shows the proportion of the population in each country reporting good and very good health, in response to the question “*how is your health in general?*” The accompanying article notes that although questions on self-reported health status “may be sensitive to differences in language and ‘culture’ between Member States, it seems worth noting that for instance ‘very good’ health is reported by as much as 53% of the Danish and as little as 8% of the Portuguese population” [20]. It is unlikely that differences in the true level of health, translations or measurement error, account for such large variations between Denmark and Portugal, or other countries within the European Union. Understanding “cultural” and other factors contributing to these differences are crucial before drawing conclusions about the comparative health status across populations.

Users of household interview surveys often ignore these limitations and interpret self-reported data at face value. An analysis of existing data from nationally representative household interview based surveys conducted in almost 50 countries, both developed and developing, documents that the comparability of data on the distribution and levels of severity of health across populations is limited, irrespective of the comparability of questions included within surveys [10]. Other studies from diverse countries, such as Argentina, India and Netherlands, have documented that sub-populations with lower expectations for health (such as the elderly) or less exposure to what constitutes good health (such as those with lower socio-economic levels) actually report themselves in better health in comparison to external tests or observations of health by professionals [21-23]. Such differences reduce the cross-population comparability of data.

An analogy provides another perspective on this challenge. If three different populations are asked a question such as “*Is the temperature today very hot, hot, mild, cold or very cold,*” self-reported responses may reflect the norms for temperature in each population rather than the tested level of temperature actually measured. For example, population A is found in a stable hot region, where the average daily temperature is 32 C, population B is found in a stable colder region, where the average daily temperature is 5 C, and population C is found in a less stable environment, with average daily temperatures fluctuating between 5 – 32 C.

Figure 2 shows the three potential response scales for three different populations. The endpoints of each scale have been fixed at the boiling and freezing point of water at sea level. For population C, the threshold between each response category is evenly spaced. This means that the difference between very hot and hot, hot and mild, mild and cold, and cold and very cold, are roughly equal, and cover the span from 0 C to 100 C. Whereas for population A, on average it must be hotter for an individual to respond “very hot” in comparison to an individual in populations B or C. Conversely, for population B, it must be colder for an individual to respond “very cold” in comparison to an individual in population A or C. Although responses to the question “*Is the temperature today very hot, hot, mild, cold or very cold,*” may be reliable and valid within each of these populations, the self-reported assessment of temperature using the available response categories are not comparable across populations. For example, the response “Very cold” in population A maps to “very cold”, “cold” and even “mild” within population B, and “very cold” and “cold” in population C.

**Figure 2. Lack of cross-population comparability: different scales of reference**



Although more complex and challenging than this illustration, when different populations or sub-populations are asked a series of questions concerning their rating of health on different domains, such as the following question on physical functioning, “*How much difficulty did you have in climbing stairs during the past four weeks, none, mild, moderate, severe or very severe?*,” self-reported responses may in one population may not necessarily be interpreted in another population in the same manner. The responses may reflect other aspects than the individual’s true difficulty in climbing stairs, and may also include differences in norms and expectations for levels of health. Such differences may render health survey results less comparable across groups or populations.

## **2.0 WHO common framework for measuring and reporting information on the health of populations**

The WHO common framework for measuring and reporting information on the health of populations includes three components that build on the major advances in the health status measurement field, and specifically address the key challenges facing this field, briefly identified in the first part of this paper. These three components include:

1. WHO Family of International Classifications on Health.
2. Operational systems of data collection on Health States of Populations
3. Summary Measures of Population Health.

The historical context of this framework is briefly described, followed by a review of the process to develop each component of the WHO common framework. The final section of this paper provides the text of the recommendations concerning the WHO common framework, made by the WHO Committee of Experts on Measurement and Classification for Health. This members of this committee, made up of experts from all WHO regions, have made recommendations towards developing a WHO resolution on measuring and reporting on the health of populations. A resolution that is endorsed by the WHO Executive Board and by World Health Assembly will enhance the institution's mandate to provide comparable data on the health of populations cross-nationally, in conjunction with Member States.

### **2.1 Historical Context**

The need for reliable and timely information on the health of populations as a critical input into the public policy process was recognized from the very inception of WHO. Indeed, the constitution ([Article 64](#)) specifically stated that "Each Member shall provide statistical and epidemiological reports in a manner to be determined by the Health Assembly" and under [Article 65](#), that "Each Member shall transmit upon the request of the Board such additional information pertaining to health as may be practicable".

Prior to the early 1980s, health reporting by Member States focused on three broad areas, namely:

- (1) causes of death,
- (2) new cases of infectious diseases, and
- (3) the number and distribution of health personnel and health institutions providing care.

Following the adoption of the Global Strategy for HEALTH FOR All by the world Health Assembly in 1981, a common format and framework for monitoring progress with the implementation of the HFA Strategy was developed, based on 12 global indicators. The indicators chosen for monitoring health status of populations were largely mortality-based (e.g., life expectancy at birth, infant and child mortality, maternal mortality) and only one measure of health (i.e., nutritional status) was included. Subsequently, the list of indicators was revised and extended to more than 70 specific indicators, including measures of the prevalence of 12 communicable or nutritional disorders.

With the epidemiological transition from communicable to non-communicable diseases, the measurement of the non-fatal consequences of diseases, particularly chronic diseases and injuries, has become increasingly relevant to all Members States. At the same time, relatively little attention has been given to the concepts, methods and data requirements to adequately incorporate the assessment of these "non-fatal health outcomes" into health status measurement strategies.

The Common Framework for Monitoring and Evaluating the Implementation of Health for All Strategies has been criticized for being too cumbersome and onerous for Member States. Reporting on more than 70 indicators, many of which are not the direct responsibility of the health sector in countries, has created substantial demands on limited resources typically allocated for health reporting. In order to meet the needs for data and information under the Health For All framework, countries were often required to carry out additional surveys which in many cases were poorly co-ordinated, data inefficient and resource intensive.

The need for greater emphasis on health rather than simply mortality in the Health for All Monitoring and Evaluation Framework was belatedly recognized. Furthermore, the concepts, terminology and methods to ensure cross population comparability of health status assessments were also not well developed under the HFA Framework. As a result, data from Member States on health outcomes was not comparable, neither over time in their populations, nor in comparison with other populations. This lack of comparability affected both national monitoring and evaluation of health status, as well as global monitoring requested of WHO.

Finally, the Health For All Framework, while touching on several specific aspects of the health situation, did not provide a coherent integrated framework for monitoring the health progress of populations. The list of indicators undoubtedly covered many important elements of the health policy process, but was not based on an integrated approach to defining health, nor overall health system performance. As a result, it was often difficult to understand how indicators were related to each other, which were more important, and how might one judge whether overall progress with the implementation of the HFA Strategy was being achieved.

The following section of the document briefly describes the development of a common framework for countries to measure and report health status of their populations, relevant to modern health information needs. The next section specifically addresses the process towards building a common framework, given the challenges identified in the first part of the document, the demands of Member States and WHO's constitutional mandate in this area.

## ***2.2 Three components of the WHO common framework to measure and report on the health of populations***

### **▪ WHO Family of International Classifications on Health**

As discussed, additional information on population health is needed beyond mortality or disease. The International Classification of Impairments, Disabilities and Handicaps (ICIDH) was therefore created in 1975 (WHA Resolution 43.24) to report the consequences of diseases and needs of individuals. ICIDH was used in several countries for field trial purposes and a revision process commenced in 1995 to address various issues, including the need to use ICIDH as a common language or framework to measure and report on the health of populations.

Specifically, it is important to provide a precise operational definition of health in order to facilitate comparable measurement across countries and the calculation of comparable summary measures. This precise definition for the purposes of reporting on the health of the populations of Member States will be provided by the ICIDH framework. Periodicity of revision will be such as to ensure stability of the classification system and incorporation of new knowledge

Currently, countries report mortality statistics primarily according to the International Classification of Diseases (ICD) system. It is worthwhile to highlight that the ICD was originally created in 1853 by the International Statistical Congress as a "uniform nomenclature of causes of death applicable to all countries". Since 1948, WHO has been responsible for the ICD and it has been revised five times: ICD-6 1948; ICD-7 1955; ICD-8 1965; ICD-9 1975; and ICD-10 1989.

The International Conference for the Tenth Revision of the ICD in 1989, held in Paris, recommended that "... WHO should endorse the concept of an updating process between revisions and give consideration as to how an effective updating mechanism could be put in place". This recommendation was subsequently endorsed by the Forty-third World Health Assembly (WHA 43.24). The Secretariat has worked with the network of WHO Collaborating Centres for the Classification and a process for the periodic updating of the ICD between formal revisions has now been established with the creation of a Mortality Forum, a Mortality Reference Group and an Update Reference Committee. It is planned that amendments to the classification will not be made more than every three years.

Over successive revisions, the ICD has become the international standard classification for all general epidemiological and many health management purposes. The ICD is useful to identify the

length of life span and reasons for death, but gives no further indication on health status among the living population.

A similar process for ICDH revision has been put in place over the last five years. A large network of several collaborating centres, and governmental and non-governmental organizations have taken part in the revision and field testing of three successive versions of the classification. The most recent review in September 2000 includes that by the WHO Expert Committee on Measurement and Classification for Health.

## ▪ **Operational systems of data collection on Health States of Populations:**

### ***Common survey instrument with standardized health profile***

Health policy development in countries will be greatly facilitated if comparable data on health status is available and utilized. To guide the national health policy process, it is essential for countries to monitor overall health progress over time and across different sub-groups in a comparable fashion. Standardized health measurement approaches adopted by Member States will ultimately lead to better international comparability and more reliable assessments of priority health needs.

To assist Member States to enhance the comparability of health measurement and reporting, a common survey instrument, including data collection methods and analysis approaches, are being developed. The goal is to develop and test survey modules for measuring health that achieve high cross-population comparability. These survey instruments are being crafted in congruence with the WHO Family of International Classifications and designed to yield practical information for Summary Measures of Population Health. The survey modules and mechanisms to ensure cross-population comparability could be integrated in routine data collection mechanisms in Member States and this information could be used to translate information into a common format for international comparisons.

The survey under development builds upon the experiences of other existing health status assessment instruments used in cross-national applications, as well as those gained from the development and testing of two other instruments that WHO has developed over the past 10 years [24]. These are the WHOQOL (WHO Quality of Life Instrument) and the WHODAS-II (WHO Disability Assessment Survey--II), that have been developed simultaneously in developed and lesser developed countries, in close collaboration with a broad network of collaborating centers and tested in a wide range of countries representing all of the WHO regions.

The selection of domains for inclusion (the common health profile) and the specific items are being developed through expert review and empirical testing. There are potentially three sets of domains in order to describe health:

- core domains of health that almost all people agree upon
- additional domains of health that some people consider as core domains
- other domains that are related to health and serve as good proximate measures of the experience of health – health related domains

Data collection and analyses methods will test and improve the cross-population comparability of population level information on health. Several workshops with experts reviewing the development of the survey, including the candidate list of domains, have been held during 2000, including with the WHO Committee of Experts on Measurement and Classification for Health. Additional consultations with Member States will continue to be held during the further development and testing of the instrument. The working group meetings leading up to the Ottawa Conference, and the conference itself, is part of this consultative process.

Furthermore, a pre-pilot survey was conducted in several countries during 2000. As a common instrument is developed and applied in a large number of countries, empirical data will provide insights into how to refine and improve the survey instrument. For example, the pilot phase will include nationally representative surveys. Initially, WHO is providing support for nationally

representative surveys in some 70 countries, through household interview based surveys and postal surveys, to take place during 2000-2001. This initial phase of testing may also include additional countries: national statistics offices may also test the instrument within existing national health surveys or pilots.

To ensure that an instrument remains the best possible instrument for measuring the health state of individuals, a mechanism will be established to periodically review the experience and propose modifications of the instrument.

## ▪ **Summary Measures of Population Health**

The regular assessment of population health for WHO Member States is a key component of the public policy process to improve health levels and reduce health inequalities in Member States. Population-level estimates of mortality, morbidity and health status/disability in countries by age, sex and cause, are useful for several public health purposes, ranging from the monitoring of new epidemics to progress in reducing old ones for which disease control programmes are in place. To adequately describe health patterns in 191 Member States according to age, sex and cause, a vast array of estimates need to be generated. It then becomes extremely difficult to ascertain the main findings of such a review or to compare levels of population health across countries unless the data are summarized in some fashion.

Summary measures of population health (SMPH) are measures that combine information on mortality and non-fatal health outcomes to represent population health in a single number. While summary measures have many potential uses, there are two that are particularly important for public health use: comparative judgements about the health of a population and comparative judgements about the relative magnitude of different diseases, injuries and risk factors as contributors to overall health. SMPH designed for comparative assessments need to answer the questions:

- (a) is population A healthier than population B?
- (b) has the health of population A improved between times  $t_1$  and  $t_2$ ?

SMPH are also required to be able to compare the magnitude of different diseases, injuries and risk factors as contributors to population health. These two key uses may best be served by different forms of summary measures of population health.

For reporting on the health of WHO Member States, SMPH provide a simple and useful summary of the vast array of components of population health. SMPH do not replace the more detailed reporting of data for specific aspects of health and mortality or for specific causes of health problems; rather they supplement these data by providing a summary measure that can be used to monitor trends and compare health across populations.

In the last two decades, considerable international effort has been put into the development of summary measures of population health (SMPH) that integrate information of mortality and non-fatal health outcomes and international policy interest in such indicators is increasing. The concept was first proposed in the 1960s and Disability-Free Life Expectancy (DFLE) was calculated for a number of countries during the 1980s. During the 1990s, DFLE and related measures were calculated for many countries and this year, for the first time, the World Health Organization published estimates of healthy life expectancy (DALE) for its 191 Member States [15]. Another form of SMPH, the Disability-Adjusted Life Year or DALY, has been used in the Global Burden of Disease Study [25]. DALYs are a health gap measure that reflects the gap between current population health and a normative standard.

Reflecting this rising interest in the academic and policy communities, the United States' Institute of Medicine convened a panel on summary measures and published a report that included recommendations to enhance public discussion of the ethical assumptions and value judgements, establish standards, and invest in education and training to promote use of summary measures [26]. At the end of 1999, WHO convened a conference of experts across a range of disciplines including descriptive epidemiology, public health, health economics and philosophy and ethics to discuss issues around the conceptual, technical and ethical basis for summary measures of population health. A book addressing these issues based on the papers presented at this conference is in preparation and a draft

was provided as one of the background documents [27] to the WHO Committee of Experts on Measurement and Classification for Health.

### **2.3 Recommendations by the WHO Expert Committee on Measurement and Classification for Health, September 2000.**

These recommendations for a WHO Executive Board Resolution have already been translated into all official languages of WHO, and transmitted to the Ministries of Health in all 191 Member States for consultation, comments and suggestions by 1 November 2000. This consultation process will contribute to the drafting of a resolution in view of the upcoming Executive Board to be held in January 2001.

The considerations of the WHO Committee of Experts on Measurement and Classification for Health, together with consultation with Member States over the next two months and other consultations, such as the Ottawa Conference, will result in a set of Recommendations for consideration by the Executive Board of WHO in January 2001. The Executive Board may wish to consider the adoption of a resolution for endorsement by the World Health Assembly in May 2001.

#### **Recommendations:**

*“The Executive Board, having considered the report on progress in the development and implementation of the WHO Family of International Health-related Classifications (WHO-FIC); operational systems for data collection on Health States of Populations, and Summary Measures of Population Health; may wish to recommend that the World Health Assembly:*

#### **[WHO Family of International Classifications on Health]**

*1.1 ENDORSE the International Classification of [Impairments] Disability and Health (ICIDH-2).*

*1.2. ENDORSE THAT WHO Family of International Classifications provide a valuable tool to describe and compare the health of the populations of WHO Member States: in terms of mortality and morbidity (ICD) and health states and related outcomes (ICIDH-2)*

*1.3. RECOMMEND THAT operational subsets (see Annex 1) of selected ICIDH Domains should be identified for various purposes including as the basis of a survey instrument or other data collection methods:*

*1.4. RECOMMEND THAT WHO Secretariat periodically revise ICIDH. Periodicity of revision will be such as to ensure stability of the classification system and incorporation of new knowledge.*

#### **[Operational systems of data collection on health states of populations]**

*2.1 RECOMMEND THAT the WHO Secretariat should orchestrate the development and periodic implementation of a common survey instrument for measuring health states.*

2.2 *ENDORSE THAT cross-population comparability is an essential requirement for reporting on health for WHO Member States in addition to cross-cultural applicability, reliability and validity. AND RECOMMEND THAT Member States use an explicit strategy to establish cross-population comparability which is incorporated into the common instrument design for each health domain.*

2.3. *RECOMMEND THAT The subset of core health domains selected from ICIDH-2 should be the basis of the development of a common survey instrument for measuring health states in periodic surveys in general populations [Attachment 1]; and THAT there should be a continued effort to develop a short list of domains to describe health states for various purposes.*

2.4. *RECOMMEND THAT the Survey Instrument should be periodically revised and maintained by WHO Secretariat to reflect experience in its application and scientific advance in the field of health state measurement.*

### **[Summary measures of population health]**

3.1 *ENDORSE THAT summary measures of population health (SMPH), combining information on mortality and non-fatal health outcomes, provide a valuable tool to summarise, monitor and compare the health of populations and to compare the relative magnitude of different causes of population health levels.*

*ENDORSE THAT both types of SMPH, health expectancies and health gaps, are needed. Health expectancies provide a simple summary to compare the health of populations and health gaps provide a simple basis for comparing the magnitude of different causes of population health levels.*

(a) *RECOMMEND THAT Member States use SMPH for monitoring on population health levels and for assessing the contributions of different causes.*

(b) *RECOMMEND THAT the Director General report annually on SMPH for Member States.*

3.2 *RECOMMEND THAT because distribution of health within populations is important, as well as level of health, separate summary measures should be used to report on level and distribution of health.*

3.3 *RECOMMEND THAT because health state valuations are a critical input to the reporting of both health expectancies and health gaps, health state valuations should be measured in population-representative samples in each Member State.*

3.4 *RECOMMEND THAT WHO provide detailed technical advice to Member States to assist them to develop and improve data sources on mortality, cause of death, morbidity and non-fatal health outcomes.*

3.5 *RECOMMEND THAT WHO, in consultation with Member States and appropriate expert networks, develop guidelines and standards for the calculation and reporting of summary measures of population health for purposes of international comparison.”*

*ATTACHMENT 1:*

*A subset of core health domains selected from ICIDH-2 for measuring health states in periodic surveys in general populations*

*Health Domains:*

*[ General Health]*

- 1. Vision*
- 2. Hearing*
- 3. Speech*
- 4. Digestion*
- 5. Bodily excretion*
- 6. Fertility*
- 7. Sexual activity*
- 8. Skin & disfigurement*
- 9. Breathing*
- 10. Pain\**
- 11. Affect\**
- 12. Sleep*
- 13. Energy/vitality*
- 14. Cognition\**
- 15. Communication*
- 16. Mobility\**
- 17. Dexterity*

*Health Related Domains*

- 18. Self-care\*: Including eating*
- 19. Usual activities\*: household activities; work or school activities*
- 20. Interpersonal relations*
- 21. Social functioning*

\* indicates candidate items for shorter list

## References

1. **Jamison DT, Frenk J, Knaul F** (1998). International collective action in health: objectives, functions, and rationale. *Lancet*. 351(9101):514-7, 1998 Feb 14.
2. **Katz S, Ford AB, Moskowitz RW. et al.** (1963). Studies of illness in the aged: the index of ADL: A standardized measure of biological and psychosocial function. *Journal of the American medical Association* 185:914-919.
3. **Ware JE, Gandek BL, Keller SD, the IQOLA Project Group** (1996). Evaluating instruments used cross-nationally: methods from the IQOLA project. In *Quality of life and pharmoeconomics in Clinical Trails*, second edition, Spilker B (ed). Lippencott-Raven Publishers, Philadelphia.4. (UN 1995)
5. **Kroeger A** (1988). Modules 11 – 12, Morbidity & Specific Diseases, conditions and symptoms. In *Training Modules for Household Surveys on Health and Nutrition*, Epidemiological and Statistical Methodology Unit, WHO, Geneva. Pp.105-122.
6. **Fisher G, Pappas G, Limb M** (1966). Prospects, problems and prerequisites for national health examination surveys in developing countries. *Soc Sci Med* 42(12):1639-1650.
7. **Ferrer M** (2000). *Health modules in household surveys in Latin America and the Caribbean: an analysis of recent questionnaires*. Public Policy and Health Program, Health and Human Development Division, Technical Report Series n. 72, Pan American Health Organization, Washington, D.C.
8. **Hupkens C** (1998). *Coverage of health topics by surveys in the European Union. Population and social conditions*, Eurostat Working Papers 3/1998/E/n. 10, European Commission, Luxembourg
9. **Gudex C, Lafortune G** (2000). *An inventory of health and disability-related surveys in OECD countries*. Directorate for Education, Employment, Labour and Social Affairs, OECD, Paris
10. **Sadana R, Mathers C, Lopez AD, Murray CJL, Iburg KM** (2000). *Comparative analyses of more than 50 household surveys on health status*. Geneva, World Health Organization, 2000 (GPE Discussion Paper No. 15).
11. **WHO** (1992). *International Statistical Classification of Diseases and Related Health Problems*, Tenth Revision (ICD-10), World Health Organization, Geneva.
12. **Bergner M, Rothman** (1987). Health status measures: an overview and guide for selection. *Ann Rev Public Health* 8:191-210.
13. **Ware JE, Keller SD** (1996). Interpreting general health measures. In *Quality of life and pharmoeconomics in Clinical Trails*, second edition, Spilker B (ed). Lippencott-Raven Publishers, Philadelphia.
14. **Murray CJL, Salomon J, Mathers C** (1999). A critical examination of summary measures of population health. *The Bulletin of the World Health Organization*, vol.78, no.8, 981-994
15. **WHO** (2000). *World Health Report 2000*. Geneva, World Health Organization.
16. **Mathers C, Sadana R, Salomon J, Murray CJL, Lopez AD** (2000). *Estimates of DALE for 191 countries: methods and results*. Geneva, World Health Organization, 2000 (GPE Discussion Paper No. 16).

17. **Ware JE, Kosinski M, Gandek B, Aaronson NK. et al.** (1998). The factor structure of the SF-36 Health Survey in 10 countries: results from the IQOLA Project. *J Clin Epidemiology* 51(11):1159-1166
18. **Nossikov A, Gudex C** (2000). *Overview of the implementation of EUROHIS*. Paper presented at the ECE-WHO Joint Meeting on Health Statistics, Ottawa, 23-25 October 2000.
19. **Robine JM, Jagger C, Egidi V** (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 Montpellier (France), Euro-REVES
20. **Eurostat** (1997). Self-reported health in the European Community. *Statistics in Focus, Population and social conditions*. ISSN 1024-4352
21. **Kroeger A, Zurita A, Perez-Samaniego C, Berg H** (1988). Illness perception and use of health services in north-east Argentina. *Health Policy and Planning* 3: 141-151.
22. **Murray CJL** (1996). Epidemiology and morbidity transitions in India. In *Health, Poverty and Development in India*, eds. DasGupta M, Chen LC and Krishnan TN, Oxford University Press, Delhi, 122-147.
23. **Mackenbach J.P, Looman CWN, van der Meer JBW** (1996). Differences in the Misreporting of Chronic Conditions, by Level of Education: The Effect on Inequalities in Prevalence Rates. *American Journal of Public Health* 86(5):706-711
24. **WHOQOL Group** (1998). Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychological Medicine* 28(3):551-8
25. **Murray CJL, Lopez AD. eds.** (1996). *The Global Burden of Disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2030*, Global Burden of Disease and Injury Series, Vol.1, Harvard University Press, Cambridge
26. **Field MJ, Gold GM, eds.** (1998). *Summarizing Population Health: Directions for the Development and Application of Population Metrics*. Institute of Medicine, Washington, D.C. National Academy Press.
27. **Murray CJL, Salomon J, Mathers CD, Lopez A, Lozano R** (2000). Summary Measures of Population Health. Geneva, World Health Organization (draft).