Mental health measurement in comprehensive national health surveys

Ulla-Sisko Lehto-Järnstedt, Arpo Aromaa

National Public Health Institute, KTL

HEALTH SURVEYS IN THE EU:

HIS AND HIS/HES EVALUATIONS AND MODELS,

Phase 2 / Subproject 2

Contents

PART II

5 MENTAL HEALTH MEASUREMENT IN NATIONAL HEALTH INTERVIEW (HIS) AND HEALTH EXAMINATION (HES) SURVEYS IN EU/EFTA COUNTRIES

ABSTRACT

Health consists of deeply interdependent physical, psychological, and social dimensions. Positive mental health is a health resource, whereas negative mental health is usually understood as mental disorders and symptoms. Validated instruments addressing mental health issues may be either psychometric scales or diagnostic measures. Psychometric measures provide the distribution of a domain in the population whereas diagnostic measures provide the diagnosis of various disorders according to current international psychiatric classifications.

We review mental health and quality of life measurement in the national comprehensive health interview (HIS) and health examination (HES) surveys. Specific psychiatric mental health surveys are not included. This report 1) reviews methodological aspects on measurement of mental health and quality of life, 2) describes and evaluates the current use of validated mental health and quality of life measures in major national HISs and HESs in EU and EFTA countries, and 3) produces tools for developing the health survey (HIS and HIS/HES) database, and 4) recommendations on mental health and quality of life measurement in national HIS and HES surveys. The final aim of this work is to improve the international comparability of results on mental health related issues. This report is part of the project 'Health Surveys in the EU: health interview (HIS) and health examination (HES) evaluations and models'.

The initial search on the use of mental health and quality of life instruments in national health surveys was carried out from the HIS/HES database year 2001 version, and completed from the June 2003 website version. In the 2003 version concerning the years 1998 - 2002, there were a total of 64 surveys, 11 with a HES component and one plain HES. Twenty-one surveys (33%) contained validated mental health or quality of life measures, two-thirds did not include validated measurement of the area. In spite of the general mental health and health-related quality of life, targets of measurement varied, and instruments addressing the targets differed substantially. The number of these measures ranged from 1 to 8 per survey. Most surveys applied only one measure, usually either a general mental health scale or a quality of life scale. There were only two widely used measures, namely GHQ-12 general mental health questionnaire and SF-36 quality of life questionnaire. Two surveys included a large number of mental health or QOL measures. In HESs, the use of these measures was much broader than in HISs.

It was unfortunate to discover that in the current comprehensive health surveys in Europe the measurement of mental health issues with validated instruments was rare, the targets of measurement varied greatly, and methods addressing the targets differed. Thus, currently these surveys offer few data for international comparison. There is a considerable need to develop the quality and comparability of measurement in the field of mental health and quality of life aspects in national HIS and HES in EU.

1 INTRODUCTION

1.1 Concepts of mental health

Health consists of closely interwoven and deeply interdependent physical, psychological, and social dimensions (McDowell & Newell, 1996; WHO, 2001). Mental health is one part of the entity of general health, and, furthermore, it emerges as a key determinant of overall health (WHO, 2001). Positive mental health is a value in itself and may be considered a health resource. It includes concepts related to personality traits, like self-esteem, sense of mastery, sense of coherence and self-efficacy. Also health-related quality of life (QOL) may be counted into positive mental health. While positive mental health is important, there is no consensus on how to promote it among people. Negative mental health is usually indicated and evaluated by mental disorders, symptoms, and problems. Mental disorders develop in the interaction of biological, psychological and social factors. The newest reports on the subject add behavioural aspects to the disorders and use the concept 'mental and behavioural disorders' (WHO, 2001).

"Mental health has been defined in various ways by scholars from different cultures. Concepts of mental health include aspects like subjective well-being, perceived self-efficacy, autonomy, competence, intergenerational dependence, and self-actualization of one's intellectual and emotional potential, among others (WHO, 2001)". It has been claimed that, from a cross-cultural perspective, it is nearly impossible to define mental health comprehensively. It is, however, generally agreed that mental health is a broader concept than lack of mental disorders. Work is often focused on mental disorders rather than mental health. The disorders are understood as clinically significant conditions with alterations in thinking, mood (emotions) or behaviour associated with personal distress and/or impaired functioning. Mental disorders are conditions not within the "normal" range (psychological), but clearly abnormal or pathological phenomena (psychiatric) (WHO, 2001).

Although mental health is crucial to the overall well-being of individuals, societies and countries [1], mental health and mental disorders are often not seen as being as

---

[1] Mental and behavioral disorders are estimated to account for 12% of the global burden of disease worldwide and more than 25% of all people suffer from a mental disorder at some time during their lives (WHO, 2001).

important as physical health (WHO, 2001). Mental and behavioral disorders affect all people (with all ages, both genders, all income groups, and both urban and rural environments from all countries and societies). Mental ill-health is a major public health concern in Europe. Valid and representative general population epidemiological data on mental disorders are needed as a foundation for public health initiatives.

1.2 Measurement of mental health

The variety of measurable mental health aspects is wide, ranging from very severe disorders to psychological well-being, adjustment, and quality of life issues that may not be identified by a medical diagnosis (cases versus non-cases) but a health resource which is normally distributed in the population. Often mental ill-health and disorders are the main concern in health surveys. Their evaluation follows usually the current international psychiatric diagnostic classifications of mental disorders. Psychological distress is a non-specific syndrome, which may include anxiety, depression, anger, irritability, and other concepts previously classified under neuroses. Depression and anxiety are core components of the distress syndrome. Several syndromes, psychiatric diagnoses, and conditions have been assessed in health surveys. Depression (affective disorders) is the most commonly evaluated single mental disorder in health surveys; it is almost always included in health surveys containing mental health aspects. Measurement targets have also included, e.g., psychotic disorders (schizophrenic and other), anxiety disorders, posttraumatic stress disorders, and personality disorders. In studies, also eating disorders, substance use disorders (alcohol, drugs, tobacco), and mental status deviance and psychogeriatric disorders (dementia) have been included in mental disorders. Measurement of the last three groups will not be reviewed here. Nor do we evaluate measurement of functional capacities and limitations. Furthermore, we do not include measurements utilizing single item questions.

Figure 1.2 Targets of measurement of mental health and mental disorders
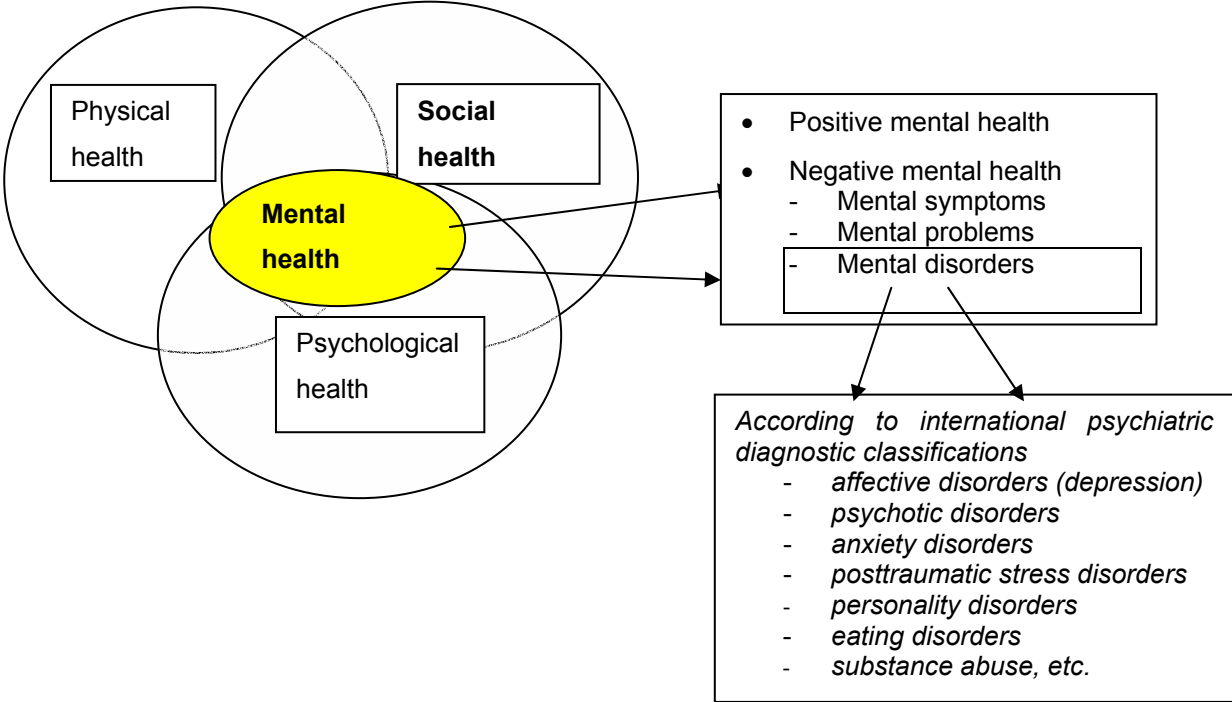
There is a wide variety of measurement techniques. There has recently been progress in consolidating the field of health measurement (McDowell & Newell, 1996). From the methodological point of view, there are two main groups of measures on mental health issues, i.e., psychometric measures and diagnostic measures. Psychometric instruments produce numerical values for different domains whereas diagnostic measures provide the diagnosis per se. Psychometric measures assess also (non-diagnostic) domains with a normal distribution in the population. Nowadays there are good methods, and they are being used more consistently in a growing number of studies. Thus, they also provide genuinely comparative information.

1.3 The ongoing EU project

An ongoing EU Health Monitoring Program (HMP) project 'Health Surveys in the EU: HIS and HIS/HES evaluations and models' collects data on health interview surveys (HIS) and health interview/examination surveys (HIS/HES) and develops methods and recommendations for them. Phase 1 of the project collected and reviewed the surveys and created a database (HIS/HES database, 2000; Hupkens & Swinkels, 2001; Koponen & Aromaa, 2001). Phase 2 of the project was initiated early in the year 2001 and it's aims are 1) to select, evaluate and recommend methods for use in health surveys and 2) to maintain and develop the health survey database. This mental health project is one of the subprojects in phase 2 and concerns evaluation of the mental health measures used in HIS/HES surveys and recommendations on measurement of mental health issues. It contributes to the first aim of the main project.

The mental health measures used in health surveys, and information on design, purposes and implementation in national surveys in European countries have been collected and are available in the database (HIS/HES database, 2000, 2001). The possibility to compare improves the usefulness of the data and it will also enhance the quality and comparability of future surveys.

2 AIMS OF THE PROJECT AND THIS REPORT

2.1 Aim of the project

This project describes the use of and evaluates the qualities and comparability of the measurement of mental health issues **in national comprehensive health interview (HIS) and health examination (HES) surveys in EU and EFTA** countries.
The specific aims are:
1. Description of the use of mental health measurement in health surveys
   - Which measures have been used (in which form, parts of measures, adapted versions),
   - what have been the research targets (substance of measurement),
   - in which surveys (the surveys, number of the surveys, and countries with these surveys), and
   - why (reason for applying certain measures).

2. Evaluation of characteristics of the measurement and measures
    - Level of validation of the measures
        - psychometric properties
        - international *versus* national (one culture only) validation
        - recommended area for application (clinical/ screening/ survey instrument)
    - Use of adapted *versus* original versions of the measures (effect on validity and comparability) and application of the measures (field work)

3. Giving recommendations on the <u>principles of proper measurement</u> of mental health issues in national health surveys, in order to enhance their international comparability.
    - Give recommendations on the type of measurement and measures suitable for use in national HIS and HES surveys
    - Include the recommendations in the HIS/HES database

2.2 Aims of this report

This report deals with methodological aspects of measurement of mental health and quality of life (Chapters 3 and 4). It describes the current use of mental health and quality of life instruments in national comprehensive health surveys (HIS and HIS/HES), and evaluates them from the view point of international comparability and development of the measurements in the area (Chapter 5).

The report has two parts. First, the most important aspects of measurement, methodology and measures of mental health issues will be reviewed and summarized. Selected major mental health measures will be described and reviewed from the point of view of survey settings, and the validity of these instruments in the population is described. Second, the use of mental health related and quality of life measures in HIS and HES surveys in EU and EFTA countries will be described and analyzed. Finally, the report contains an evaluation of and recommendations on the measurement of mental health and quality of life in health surveys.

PART I

3 MEASUREMENT OF MENTAL HEALTH

'Mental health' as a measurement target in health surveys can be defined in different ways. Mental health related measurement targets range from severe disorders to positive mental health issues, such as well-being and quality of life. The latter cannot be identified by a diagnosis (cases versus non-cases) but they follow a normal distribution in the population. There are two main groups of measures addressing mental health issues, arising partly from different scientific traditions, namely psychometric measures and diagnostic measures. Psychometric measures can address both diagnosable disorders and well-being issues. These instruments produce numerical values for different domains of the construct. Psychometric measures have a long tradition in psychology and they are used for evaluating various topics, including health, mental functioning and mental disorders. Diagnostic measures provide the diagnosis per se to differentiate the healthy population from people with the disorder. They usually follow the international psychiatric diagnostic classifications; each mental disorder is defined in the classification by the existence of specific clusters of symptoms. Diagnostic measures are psychiatric tools, and their research tradition arises from clinical practice in diagnosing and treating mental disorders.

The health measures may also be grouped into instruments on 1. social health, 2. psychological well-being and morbidity, including depression, 3. mental status testing, 4. pain measurement, and 5. quality of life (McDowell & Newell, 1996). Psychological well-being, and morbidity, evaluated as mental disorders indicating mental ill-health, is the main interest in this report. Depression measurement is most often - and most appropriately - handled as an entity of its own, especially because depression measures are common and usually possess high validity. Social health is rarely measured in European health surveys, although several methods have been designed (McDowell & Newell, 1996). Quality of life measures represent measurement of positive mental health and they are included because of their importance for health and the wide spread use of quality of life measures in health surveys.

We will not discuss mental status measures, pain measurement or single-item health measures. 'Mental status' measures evaluate the cognitive status of a person: cognitive function, cognitive impairment, and dementia (McDowell & Newell, 1996). Pain measurement is peripheral from the point of view of mental health and mental disorders. The single-item 'perceived health' questions, some times also interpreted as quality of life measures, as well as other measurement consisting of one single item are also not in the core of the current discussion.

3.1 Psychometric measures

"Psychometric testing sums up performance in numbers" (Cronbach, 1990). The respondents are usually asked to indicate the presence, frequency or intensity of a behavior, experience, perception, feeling, symptom, etc.. The responses to the individual questions are aggregated to form scales, measuring a particular subconstruct of the concept. These instruments produce several numerical values for different domains of the construct or a single score of the construct. The psychometric methods give numerical values, usually on an ascending scale (e.g., from 0 to 3) to certain items or questions representing a certain domain, and can therefore be analyzed by various statistical methods.

Psychometric measures have considerable advantages in the measurement of perceived health, or perceived mental health issues, in health surveys. By psychometric methods it is possible to collect accurate data with reasonable resources from large population based samples. Furthermore, validated measures allow the comparison of the results across samples from different populations. In health research, several types of measurements have been used to evaluate perceived health, well-being, and mental health. As always, the documentation of the measurement and methods used is essential. The type of measurement and the measures used – along with information on the validation process of the measures – should be reported.

In the evaluation of the psychometric methodology used, some key issues presented below should be considered.

3.1.1 Types of methods and their application

Self-report *versus* expert-report scales

The psychometric measures may be self-administered scales or meant to be filled-in by a professional expert, e.g., a clinician. A great deal of evidence suggests that self-report and expert-report scales have a roughly similar amount of strengths and weaknesses (Derogatis & DellaPietra, 1994). Both approaches have advantages in measuring mental health, but neither approach can be said to be more effective in detecting psychiatric disorders.

Self-report measures tend to be brief, inexpensive, and are tolerated well by the examinees; the cost-benefit relation is good (Derogatis & DellaPietra, 1994). The use of clinician rating measures in population surveys requires a lot of resources. Self-report scales are transportable; they may be used in several settings and they minimize professional time and effort: the administration, scoring, and evaluation require little or no professional input. Perhaps the biggest advantage of self-report scales is that the measure is completed by the only person experiencing the phenomena – the respondent.

The presence of an expert interviewer is sometimes seen to guarantee objective data, whereas self-report scales may be seen as subjective. This feature of self-report measures may also include the greatest potential source of error (patient bias in reporting) (Derogatis & DellaPietra, 1994). Because the respondents provide the data, it is possible that they consciously or unconsciously bias the responses. Anyhow, empirical studies have indicated that these distortions become a problem only when there is a possibility of obvious personal gain. Furthermore, data collected by an interviewer may be influenced both by the interviewee and the interviewer.

Probably the greatest limitation of self-report scales is their inflexibility, that is, the questions can not be altered or modified according to the respondent's previous answers. Second, only the subjects' written responses are available, and no additional data on reactions of the respondent is available. The choice between self-

report and interview must be based on the purpose and setting of the measurement and the reasons why a certain alternative has been chosen should be reported.

The way in which an instrument has been used in a survey should be reported so that those using the results have sufficient information about the data. The manual of a measure is usually the principal source of information about the application of an instrument. However, several measures can be applied in several ways, e.g., as a self-report scale or as part of an interview.

3.1.2 What is being measured, the issue of validity

The validity of an instrument refers to the extent to which the instrument or a score measures what it is supposed to measure: whether it has the intended interpretation (Cronbach, 1990; Derogatis & DellaPietra, 1994; Gandek, Ware, & the IQOLA Project Group, 1998). Assessment of validity of a measurement involves a comparison of the instrument with a standard. However, it is not always clear what should be used as the standard. The validity of a measure in the health field has most often been evaluated by means of *content*, *construct*, and *criterion validity* (Cronbach, 1990; Gandek et al., 1998). These are not distinct lines in investigating validity; validation ought to integrate all of these aspects.

Content validity examines the extent to which a measure assesses those characteristics and concepts it was intended to measure; that is, whether the measure offers an adequate sample of the content of a construct (Cronbach, 1990;Gandek et al., 1998). To evaluate content validity requires something against which to compare the content of the measure. These standards can be well-accepted theoretical definitions, published standards, or interviews with those experiencing the types of health problems under study (Gandek et al., 1998).

Construct validity is a process in which the construct of a measure is being investigated; a way to organize what has been observed (Cronbach, 1990). Statistical methods (factor analysis, mean scale score comparison across groups, correlations, etc.) are applied to evaluate construct validity. *Convergent validity* and *discriminant validity* are subtypes of construct validity (Gandek et al., 1998). To

investigate convergence of an indicator two or more kinds of data are collected to provide evidence on the construct measured (Cronbach, 1990). If these indicators agree, the proposed theoretical interpretation is supported, i.e. the measures are measuring the same concept(s). This is called the multi-method principle; convergent validity indicates whether different measures of the same construct provide similar results. Second, discriminant validity shows whether the scores obtained are separable (Cronbach, 1990; Gandek et al., 1998). Scores supposed to measure distinct subconcepts should not correlate too highly; a measure said to evaluate a certain subconcept should not correlate highly with a measure that is said to measure something else. Discriminant validity indicates whether a measure of one underlying construct can be differentiated from another construct.

Criterion validity evaluates whether the scores obtained are systematically related to one or more criteria by using measures and data (the criteria) which are independent of the scale in question (Gandek et al., 1998). The hardest part of criterion-oriented validation is obtaining satisfactory criterion data (Cronbach, 1990). There are some well-known widely used measures which have become to be used as criteria against which the quality of measurements is compared. Criteria may include other measures and data gathered concurrently (concurrent validity) or after some time interval (predictive validity) (Gandek et al., 1998). The predictive validity of a health status measure can be tested by predicting some relevant future event (e.g., incident disease, death or utilization of health care services).

3.1.3 Setting standards for health measures

A standardized measure is a validated scale with norms. There is a cut-off point which may be applied to judge which values of the score should lead to certain conclusions. In the absence of agreed-upon criteria, or "gold standards" for validating health measures, normative data can be useful in interpreting scale scores (Gandek et al., 1998). The use of normative data enables the interpretation of the scale score for an individual respondent (or the average score for a group of respondents) in comparison to the distribution of scores for other individuals in the norming sample. Norm-based comparisons require proper norms derived from a well-defined and

representative sample of the population of interest. In population surveys there is a need for normative data based on findings on population-wide samples.

Some international projects have been carried out to validate mental health measures internationally. An International consortium (ICPE) has been created for facilitating cross-national research using the CIDI measure (Kessler, 1999). The World Health Organization Quality of Life Group (The WHOQOL Group, 1995, 1998b; WHOQOL Group, 1994) developed a health-related quality of life measure collaboratively in 15 culturally diverse centres. The International Quality of Life Assessment (IQOLA) Project was established to translate and validate the SF-36 Health Survey measure in several countries throughout the world (Gandek et al., 1998; "Journal of Clinical Epidemiology, Special Issue on SF-36 measure," 1998; Ware, Gandek, & for the IOQLA Project Group, 1998). The project has produced a norming protocol for the collection of general population normative data (Gandek et al., 1998). It has three broad stages: translation, tests of scaling assumptions, and empirical validation and norming. Translations of measures should follow a standard protocol, including multiple forward and backward translations and qualitative and quantitative evaluation of the translations. Formal psychometric tests of the assumptions underlying item scoring and construction of multi-item scales should be conducted. Finally, data from clinical trials, general population surveys, and other studies should be analyzed to address issues of validity and evaluate the equivalence of interpretations across countries.

3.1.4 Repeatability of the measurement, the issue of reliability

Reliability, or consistency, is concerned with error in measurement (McDowell & Newell, 1996). This refers to the consistency or stability of the measurement process across time, patients, or observers. All scientific measurement is based on consistency or replicability; reliability concerns the degree of replicability inherent in measurement. Reliability refers to the question: if the measurement is repeated (twice or more), will the score reports agree and how closely. To what degree do two clinicians agree on a psychiatric rating scale? (Cronbach, 1990; Derogatis & DellaPietra, 1994). Reliability can be thought to be the converse of measurement error. It represents the proportion of variation in measurement due to true variation in

the attribute under study. In practice, observed estimates of reliability depend both on true variation and on analytic measurement errors. The latter may contain both random and systematic components. It is sometimes possible to separate the component of reliability due to the measurement and the component due to the inherent nature of the phenomena to be investigated.

3.1.5 Selecting and evaluating measures

The first step in any study is to decide what actually should be evaluated or measured, i.e., to clarify the measurement target. The second step is to analyze the construct or constructs which are chosen to be the target of measurement. At this point the constructs and sets of constructs behind the planned measurement should be explicitly defined. In principle, the definition of the construct should guide the selection of the measure not *vice versa.* When a potentially suitable measure (or measures) has been identified, the validity of the measure should be reviewed and evaluated. The literature on psychometric properties of a measure, e.g., on the validity of the instrument, provides information on the pros and cons of the measure and this information must be compared to the measurement target chosen. When an entirely new measure is being used, or an old one is used in a new context (in different population, culture, language, etc.), the empirically obtained results by the measure should be investigated for their validity.

Lee J Cronbach (Cronbach, 1990) presents a form which may be used as a checklist to evaluate quality and validity of a psychometric measure. According to him one should check, e.g., the following aspects when choosing a psychometric measure:
1. Groups to which applicable
2. Practical features
3. General type
4. Costs
5. Scoring services available
6. Time required
7. Purpose for which evaluated
8. Author's basis for selecting items
9. Adequacy of directions; training required to administer
10. Comments regarding design of measure
11. Validation against criteria: number and type of cases, criterion measure, time interval, result
12. Other empirical evidence indicating what the measure measures

13. Comments regarding fairness
14. Comments regarding validity for particular purposes
15. Generalizability (procedure, cases, results)
16. Long-term stability (procedure, time-interval, cases, result)
17. Norms (type of scale, selection of sample)
18. Comments regarding adequacy of above for particular purposes
19. References
(Cronbach, 1990).

3.2 Diagnostic mental health measures

Clinical interviews as a procedure for obtaining diagnoses on mental disorders do not have perfect reliability and validity. In fact, it was a lack of faith in the clinical interview (the work of a clinician) which instigated the move to develop structured diagnostic interviews (Robins, 1989). Diagnostic mental health measures differ from other mental health status measures by generating diagnoses of mental disorders according to psychiatric definitions and criteria. There are relatively few diagnostic mental health measures and they always contain an interview either by a clinician or by a trained non-clinician.

The diagnostic measures map the symptoms most often onto DSM-IV criteria (Diagnostic and Statistical Manual of Mental Disorders, the number at the end of the abbreviation referring to the latest edition, now 4th) and ICD-10 (International Classification of Diseases) criteria and report whether the criteria are satisfied. Perhaps the most valued diagnostic tool of mental health is the *Structured Clinical Interview for DSM (SCID)*, a semi-structured interview for making psychiatric diagnosis according to the DSM classification (Spitzer, Williams, Gibbon, & First, 1992). It is administered by a clinician or trained mental health professional familiar with the DSM classification and its diagnostic criteria. Another tool for specialist interviewers is the *Schedules for Clinical Assessment in Neuropsychiatry (SCAN)* by WHO. The *Composite International Diagnostic Interview (CIDI)* is a measure package also by WHO. CIDI is a fully structured standardized diagnostic tool for trained non-clinical interviewers. It maps the symptoms onto DSM-IV and ICD-10 diagnostic criteria and reports whether the criteria are satisfied. CIDI was designed for use in large-scale community epidemiological surveys and other settings without clinicians to make diagnoses (Cottler et al., 1997; Kessler, 1999). It enables diagnoses to be computer-generated according to the criteria.

The diagnostic mental health questionnaires are tightly tied to the diagnostic criteria used (Robins, 1989), and their purpose is to identify 'healthy' and 'ill' in order to determine the prevalence of the disorder. In population based health surveys, it is often useful to separate 'cases' from 'non-cases' with the highest possible sensitivity and specificity. As a consequence, diagnostic mental health measures usually do not produce numerical values on all subjects but divide them into two (cases, non-cases) or three (definite, probable, non-cases).

3.3 Comparability of measurements across studies

The possibility to compare between different studies enhances the usefulness of the data. In particular, health policy benefits from the understanding of the relative health status in one country compared to others. Currently, the comparability of findings of national health surveys leaves much to be desired. Problems arise from the variety of different mental health measures used (HIS/HES database, 2000, 2001, 2002). In a recent review, a total of 60 studies including mental health indicators were reported from 12 European countries (Hibbett, Jagger, Polge, Cambois, & Ritchie, 1999). Sixty-seven measures were identified, out of which 70% were used in a single country only. Anyhow, half of the studies were found to concern the over-65 age group. Most commonly used measures were actually tests of cognitive status [2]. Depression and dementia were consistently measured because they were considered common mental health priorities.

The validity of the instruments has often not been thoroughly assessed nor reported. Not all the measures have been validated at population or international level. It cannot be assumed that a questionnaire, however extensively tested in the originating country, will be valid and reliable once it has been translated. Thus the validity of a measure should be evaluated in every culture and language in which it is used.

---

2 The most used measures were the Mini Mental State Examination (MMSE), the Geriatric Mental State Schedule (GMS), and the Cambridge Mental Disorders of the Elderly Examination (CAMDEX). MMSE is a standard tool for cognitive assessment, and GMS and CAMDEX geriatric measures.

This step need not be repeated, if a measure is originally internationally developed and validated. There are now a handful of internationally developed and validated measures. With these measures, it is possible to carry out research collaboratively in different cultural settings, and to directly compare the results obtained. Use of internationally validated measures improves comparability. It may be best to use measures developed and validated in international settings, or, if this is not possible, to use measures for which internationally acceptable standards exist.

In section 3 of this report we evaluate some mental health related measures used in health surveys. In section 4, the mental health measures used in health surveys in EU/EFTA countries are listed.

4 MENTAL HEALTH MEASUREMENT IN HEALTH SURVEY SETTINGS

Both mental health and physical health are important for the overall well-being of individuals, and they are also equally important from the public health point of view. Mental health measures have not been included in national health surveys until recently, due perhaps to a lack of agreement on a general definition of mental health as well as to problems of finding measures of sufficient brevity (http://www.un.org/Depts/unsd/disability/methods/ac.81-7-5.pdf). However, epidemiological studies have provided plenty of research findings on incidence, prevalence, determinants and consequences of poor mental health. The measurement of physical disorders tends still to play the main part in most health surveys. This is unfortunate since it may lead to under-estimation of the role of mental health in overall health and well-being. On the other hand, specific mental health surveys provide a limited view of overall health.

The methods used to evaluate mental health issues have varied substantially (HIS/HES database, 2000, 2001, 2002). Also the extent of validation of the measures has been very variable. There are, however, many carefully validated measures developed to evaluate mental health in population samples (McDowell & Newell, 1996). On the other hand, several scales have been developed and used in a single study only and the validation process – or it's reporting – has been insufficient. Some validated psychometric measures are frequently employed both in population based

surveys and as psychiatric screening instruments (Derogatis & DellaPietra, 1994; McDowell & Newell, 1996). In addition, there is at least one relatively new diagnostic measure which is internationally developed for survey use (Kessler, 1999).

4.1 International projects to foster cross-cultural comparative research on mental health and quality of life

REVES is an International Network on Health Expectancy set up in 1989 by health institutes from France, Canada and USA (http://www.prw.le.ac.uk/, http://www.prw.le.ac.uk/cgi-bin/reves). The *Euro-REVES project*, a European Concerted Action on Harmonization of Health Expectancy Calculations, was set up in 1994 and it has a mental health subcommittee: Euro-REVES Mental Health Group. It also created a mental health database (Sigmund) on health surveys and general population epidemiological studies in Europe including a mental health component, but this database is not being updated anymore.

Euro-REVES, EuroHIS and *HMP Mental Health Group* have recently tried to reach consensus on the recommended measures for mental health in surveys (Jagger & Robine, 2001, also in: www.un.org/Depts/unsd/disability/methods/ac.81-7-5.pdf). The measures considered most suitable by the various projects were different. Some consensus has been reached on measuring 'psychological distress' (MHI-5 from SF-36 if also CIDI-SF, otherwise GHQ-12), 'positive mental health/ psychological well-being' (SF-36 energy/ vitality items + Andrews item on happiness), 'anxiety and depression' (CIDI-SF) and also on some indicators, e.g., on substance use disorders, suicide attempts, dementia, and cognitive functioning. There was also agreement on the usefulness of GHQ-12, SF-36 (as parts), and CIDI-SF (or parts).

The World Mental Health 2000 (WMH2000) is a WHO project to coordinate the implementation and analysis of general population epidemiologic surveys in all WHO Regions (see, http://www.hcp.med.harvard.edu/icpe/WMH2000.html). The aim of WMH2000 is to obtain accurate cross-national information about the prevalence and correlates of mental, substance use, and behavioral disorders. It carries out a series of parallel community epidemiological surveys in countries throughout the world including a third series of CIDI surveys, 100 000 interviews in 10 countries.

WMH2000 collaborators are coordinated through the WHO International Consortium in Psychiatric Epidemiology (ICPE).

The International Consortium in Psychiatric Epidemiology (ICPE) is a consortium funded by the US National Institute of Health (Kessler, 1999) and (http://www.hcp.med.harvard.edu/icpe/index.htm), and it was founded in 1997 by WHO. ICPE is designed to facilitate cross-national comparative epidemiologic studies on psychiatric disorders. This is done by bringing together researchers from around the world with data from epidemiological surveys based on the CIDI measure to collaborate in comparative analyses.

The World Health Organization Quality of Life Group (The WHOQOL Group, 1995, 1998b; WHOQOL Group, 1994) developed a health-related quality of life measure (WHOQOL) collaboratively in 15 culturally diverse centres, including four EU countries. They address WHOQOL to epidemiological multi-centre studies to permit research questions, which cannot be studied in single site studies, e.g. the impact of social and cultural factors on the effect of diseases on quality of life.

The International Quality of Life Assessment (IQOLA) Project was established in 1991, in order to translate the Short-Form Health Survey Index of Quality of Life (SF-36), and to validate, norm, and document the translations as required for their use in international studies (Bullinger et al., 1998; Gandek et al., 1998; Sullivan, Karsson, & Ware, 1995; Ware et al., 1998 and in a special issue of J of Clin Epidemiol 41, 1998, no. 11). SF-36 has now tens of validated translations available in different languages. For survey settings, an abbreviated version of SF-12 has been developed, and a still shorter version SF-8 is under development.

The Health Monitoring Programme (HMP) of the European Commission has the aim to contribute to the establishment of a Community Health Monitoring System. *The Indicators for Mental Health Monitoring in Europe (HMP Mental Health)* project, funded by the HMP, has recently outlined a proposal for mental health indicators to monitor mental health in Europe*: Establishment of a set of mental health indicators for European Union* (Lehtinen, 2002; Lehtinen et al., 2001). The work was based on the division of health indicators into four main areas proposed by the *Integrated approach to establishing European Community Health Indicator (ECHI) project (ECHI*

Project, 2001) [3]. The HMP Mental Health project proposed the following list for a set of mental health indicators.

From the ECHI category 2 **Health Status**:

- Cause specific mortality (suicide; harmful events, intention unclear; drug related deaths; PYLL fraction: suicide),
- Disease specific morbidity (generalised anxiety disorder; major depression; alcohol dependency; suicide attempts), and
- Generic morbidity (psychological distress; psychological well-being a) energy, vitality in SF-36, b) Andrews single item question on happiness; role limitations due to emotional problems).

From the ECHI category 3 **Determinants of health**:

- Personal conditions (sense of mastery; optimism) and
- Social and cultural environment (social support; social isolation; social networks; life events).

From the ECHI category 4 **Health Systems**:

- Prevention, health protection and promotion (suicide prevention projects; projects to support parenting skill),
- Health care resources (psychiatric beds; psychiatrists; child psychiatrists; other professionals in mental health),
- Mental health care utilisation (nr of inpatient episodes; nr of inpatient episodes for minors; nr of long stay patients; use of outpatient services; self-reported use of mental health services; consumption of psyvhotropic drugs; nr of disability pensions; money spent on disability; sickness compensation periods),
- Expenditure (total national expenditure, proportionate national expenditure, proportionate national expenditure for minors), and
- Health care quality indicators (availability of national quality accreditation).

---

[3] The four main categories (and their subcategories) of the ECHI list were 1. Demographic and socioeconomic factors (Population; Socio-ecomomic factors), 2. Health status (Mortality; Morbidity, disease specific; Generic health status; Composite health status measures), 3. Determinants of health (Personal factors; Health behaviours; Living and working conditions), 4. Health Systems (Prevention, health protection and health promotion; Health care resources, Health care utilisation; Health expenditures and financing; Health care quality/performance).

4.2 Diagnostic mental health measures: measurements with the CIDI package

The Composite International Diagnostic Interview (CIDI) (WHO, 1997; Wittchen, 1994) seems to be the most carefully validated instrument of psychiatric epidemiology. CIDI is a fully structured standardized diagnostic tool for trained non-clinical interviewers. Recent versions map the symptoms into DSM-IV and ICD-10 diagnostic criteria and report whether the criteria are satisfied. Standardized probing of CIDI is conducted to determine if a symptom meets the severity treshold, and if that symptom is attributed to physical illness, medicine, or drugs or alcohol. CIDI covers eating disorders, dementia, substance use disorders (alcohol, drugs, tobacco), schizophrenic disorders, affective disorders (several), anxiety disorders, posttraumatic stress disorders, somatization disorder, psychosexual disorders, pathologic gambling, and antisocial personality disorder. CIDI was designed for use in large-scale community epidemiological surveys and other settings in which it is not feasible to have clinicians make diagnoses (Cottler et al., 1997; Kessler, 1999). CIDI has also been designed to foster cross-cultural comparative research (Kessler, 1999). It can be used to measure both life-time disorders and past year/current disorders.

To foster cross-cultural comparative research, a multinational Editorial Committee translated and field tested the instrument in many different countries (Kessler, 1999). ICPE promises to provide valid and representative general population epidemiological data on psychiatric disorders needed for public health initiatives and, eventually, to create a multinational databank. A huge amount of validation work has already been done with the CIDI instrument with hundreds of thousands of interviews from tens of countries (Andrews & Peters, 1998; Kessler, 1999; Wittchen, 1994). The inter-rater reliability has been demonstrated to be excellent, the test-retest reliability good, and the validity good, given the methodological constraints (Andrews & Peters, 1998). However, when compared to clinical diagnoses a large proportion of the CIDI diagnoses of psychosis in a community sample are false positives (Kendler, Gallagher, Abelson, & Kessler, 1996).

Domains of CIDI Short Form (CIDI-SF) (Kessler, Andrews, Mroczek, Ustun, & Wittchen, 1998), available in http://www.who.org/. have recently been recommended for monitoring mental health in HIS settings in Europe (Lehtinen et al., 2001). The

recommended parts have consisted of, e.g., generalized anxiety disorder (GAD) and depressive disorders. The CIDI-SF assesses eight syndromes (of DSM-III-R): generalized anxiety disorder, agoraphobia, panic attack, drug dependence, social phobia, simple phobia, major depressive episode, and alcohol dependence. It has been pilot tested in a national representative telephone survey and is claimed to have an average completion time of seven minutes compared to over an hour for the full CIDI.

CIDI enables diagnoses to be computer-generated according to the psychiatric diagnostic criteria. Computerized versions of CIDI have been developed at least in Australia (CIDI-Auto) (Korten & Henderson, 2000; Peters & Andrews, 1995) and in Germany (M-CIDI) (Wittchen, Lachner, Wunderlich, & Pfister, 1998). CIDI-Auto was used in a large community sample survey on mental health and well-being in Australia (Korten & Henderson, 2000). M-CIDI was developed in Munich Germany. It has been reported to be acceptable for the respondents, efficient, and having sufficient or good reliability (Wittchen et al., 1998). M-CIDI has been reported to have a shorter average administration time (63 min) than the standard WHO-CIDI, although it includes more diagnoses and subtypes.

The CIDI measures are relatively new, and until recently they have been rarely used in European countries (Hibbett et al., 1999). In WMH2000, in the European region, there is a parallel set of six nationally representative mental health surveys applying CIDI. Furthermore, in Europe at least two national health examination surveys (in Germany and in Finland) have recently employed the CIDI (HIS/HES database, 2002; The HISHES Group, 2003).

4.3 Psychometric measures on psychological well-being and psychiatric disorders

Some validated psychological measures are frequently used in population surveys and as psychiatric screening instruments (Derogatis & DellaPietra, 1994; McDowell & Newell, 1996). Especially the standardized measures with cut-off points may be directly applicable for screening purposes. Contrary to diagnostic measures psychological measures produce numerical data for all respondents.

Along the lines of previous work (Derogatis & DellaPietra, 1994; McDowell & Newell, 1996) and the HIS/HES database (HIS/HES database, 2000), two general mental health measures and five depression measures are briefly presented here. This chapter is followed by a presentation of quality of life measures. These are not measures on mental health only, but have a broader scope including health status, well-being and life quality aspects (the "positive" mental health). Some types of measures and some available instruments will be introduced. The psychometric data on the commercially available (copyright) tests, such as SCL-90 and GHQ, are available from their published administration manuals and for the non-commercial measures in published papers. Measures on depression are presented separately in a chapter of their own. In addition, the general mental health measures and quality of life measures contain items of several mental health concepts, e.g., on depression (McDowell & Newell, 1996; Schmitz, Kruse, Heckrath, Alberti, & Tress, 1999, and chapter below).

4.3.1 General mental health questionnaires

Symptom Check-List-90 (SCL-90) and Brief Symptom Inventory (BSI)

The Symptom Check List 90 (SCL-90) is a widely used psychiatric questionnaire for detection and measurement of psychopathology (Derogatis & Cleary, 1977; Derogatis & DellaPietra, 1994; Holi, Sammallahti, & Aalberg, 1998; Schmitz et al., 1999). It is a 90-item, multidimensional, self-report symptom inventory designed to measure self-reported symptom intensity on different subscales. Each of the 90 items is rated by the respondent on a 5-point scale from 0 (not at all) to 4 (extremely). The answers [4] are combined into nine primary symptom dimensions: Somatization, Obsessive-Compulsive Disorder, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism.

---

[4] Examples of items in SCL-90: 'Pains in lower back', 'Soreness of your muscles' (dimension Somatization), 'Trouble remembering things', 'Having to check and double-check what you do' (Obsessive-Compulsive), 'Feeling uneasy when people are watching or talking about you', 'Feeling that people are unfriendly or dislike you' (Interpersonal Sensitivity), 'Feeling blue', 'Feeling hopeless about the future' (Depression), 'Nervousness or shakiness inside', 'Feeling tense or keyed up' (Anxiety), . . .

SCL-90 has a long development history. It evolves from the Hopkins Symptom Checklist, which has roots both in the Discomfort Scale from the 1950s and the Cornell Medical Index from 1948 (Derogatis & Lazarus, 1994; Holi et al., 1998). Some items of these scales can be traced back to the very first self-report symptom inventory (The Woodworth Personal Data Sheet, from the year 1918). The SCL was introduced in the early 1970s. The first five of the dimensions presented above were developed from factor analytic studies, and the last four were rationally developed and later validated. SCL-90-R, introduced in early 1980s, is a revised version of SCL-90, two of the 90 items being different than in the original version. SCL-90-R norms have been developed for adult community respondents, various types of psychiatric outpatients and inpatients, adolescent non-patients, etc. (Derogatis & DellaPietra, 1994; Derogatis & Lazarus, 1994). However, the results concerning the validity of the instrument on the whole are controversial. It has been indicated that the strength of SCL-90 is its ability to screen patients from the community sample. Furthermore, SCL-90 is a center piece of a series of tests enabling assessment of clinical status via both self-report and expert clinical judgement, because there are several companion clinical rating scales to the SCL-90-R (Derogatis & Lazarus, 1994).

SCL-90-R was designed to be interpreted in terms of three distinct but related levels of information: global scores, dimension scores, and item scores, the optimal interpretation depending on integration of information from all three sources (Derogatis & Lazarus, 1994). The global scores of distress are the Global Severity Index (GSI), the Positive Symptom Distress Index (PSDI) and the Positive Symptom Total (PST). In addition, SCL-90-R is capable of producing a multidimensional symptom profile (the nine dimensions listed above), providing a syndromal context within which a score of a certain dimension can be evaluated and compared. Furthermore, some item scores are useful when interpreting the meaning of certain dimension scores. Although SCL-90 represents no doubt a significant psychological test instrument, it may be too sophisticated for use in general health survey settings.

The Brief Symptom Inventory (BSI) is the short form of the SCL-90 (Derogatis & DellaPietra, 1994; Derogatis & Melisaratos, 1983; Derogatis & Lazarus, 1994). It measures the same nine symptom dimensions using 53 items. Dimension scores of BSI correlate highly with those of SCL-90-R, and the brief score is reported to share most psychometric characteristics with the longer one. A fifty-three item scale is, however, not 'brief' for survey purposes.

Both the SCL-90-R and the BSI have been translated into tens of languages (Derogatis & DellaPietra, 1994).

General Health Questionnaire (GHQ)

The General Health Questionnaire (GHQ) was developed in the late 1970s as a 60-item, multidimensional, self-report screening instrument to detect current, diagnosable psychiatric disorder (Derogatis & DellaPietra, 1994; McDowell & Newell, 1996). It was designed for use in general population surveys, in primary medical care, or among medical outpatients. It was meant to be a "first stage" screening instrument for psychiatric illness (leaving the diagnostics to additional psychiatric interview). GHQ may be used in clinical settings or in surveys to identify potential cases. It has been developed to identify two main areas of problems: "inability to carry out one's normal 'healthy' functions" and "the appearance of new phenomena of a distressing nature", (McDowell & Newell, 1996). It focuses on breaks in normal functioning and is therefore associated with distress. GHQ is seen to measure the concept of psychological distress and it covers four elements of distress, namely depression, anxiety, social impairment, and hypochondriasis.

GHQ has 60-, 30-, 28-, 20- and 12-item versions. All items of the shorter versions are included into the longer versions; they are comparable for these parts. Items of all versions of GHQ can be found in (McDowell & Newell, 1996). GHQ-12 [5] has produced results comparable to the longer versions (Goldberg, Gater, Sartorius, & al., 1997). GHQ-12 is an instrument for the detection of non-psychotic mental disorders (Schmitz et al., 1999). Four subscales have been factor analytically

---

[5] Example items of GHQ-12: Have you recently 'been able to concentrate', 'been capable of making decisions', 'enjoyed normal activities', 'been unhappy and depressed', 'been losing confidence in yourself'.

derived: somatic symptoms, anxiety and insomnia, social dysfunction, and severe depression (Derogatis & DellaPietra, 1994). GHQ-12 has shown high reliability when used in general population samples with long intervals (Pevalin, 2000) and high validity when administered in several languages and countries (Bhui, Bhugra, & Goldberg, 2000; Goldberg et al., 1997).

GHQ has been validated for use in screening and outcome assessment in numerous populations, including many community samples (Derogatis & DellaPietra, 1994). GHQ is one of the most widely used and thoroughly validated [6] screening instruments for psychiatric disorder internationally (McDowell & Newell, 1996). The validation work has also included comparisons with numerous other scales addressed to psychological morbidity. GHQ may be the most validated and tested of the methods reviewed here. GHQ-12 has also been widely used in health surveys across Europe.

4.3.2 Depression measures

Depression scales are among the oldest, the most widely used, and also most validated mental health measures used in health surveys. There is a powerful tradition in measuring depressive symptoms and clinical depression. Most of the scales have been thoroughly validated and have been used in numerous studies in many countries (McDowell & Newell, 1996). The best depression measures seem to be 20 to 40 years old. Furthermore, there is good agreement among reviewers on a "ranking list" of leading methods for measuring depression (McDowell & Newell, 1996).

The use of the term depression also in everyday language hampers the formulation of a precise, clinical definition (McDowell & Newell, 1996). Depression covers a wide range of states from feeling sad, helpless or demoralized to a major depressive episode. 'Depression' may be used to describe:

---

[6] " The GHQ offers a leading example of how health measurement methods should be developed. It was well founded on a clear conceptual approach, the initial item selection and item analyses are fully documented, and the questions are have not been revised by subsequent users. The validation studies have been throughout and extensive . . . " (McDowell & Newell, 1996, p. 234).

1. An affect: a subjective feeling tone of short duration.
2. A mood: a state sustained over a longer period of time.
3. An emotion: comprised of the feeling tones along with objective indications.
4. A disorder: with characteristic symptom clusters, complexes, or configurations.

These form a continuum in the population, so that the frequency of normal mood fluctuations is very high and the frequency of true depressive episodes or disorders is relatively low. A depressive episode itself - the disorder - may be mild, moderate, or severe.

Most depression measures have been intended for clinical or research use. McDowell & Newell (McDowell & Newell, 1996) describe only one measure developed specially for survey use: namely The Center for Epidemiological Studies Depression Scale (CES-D), and two more instruments to be used both in screening and clinical work: Beck Depression Inventory (BDI) and Geriatric Depression Scale. Four instruments are briefly presented below. Of these four, the first two have been developed or widely used in screening for depression or in population surveys. The third has been developed for clinical use although it has also been used in survey settings. The last one is a truly clinical tool, a semi-structured clinical interview method.

The Center for Epidemiological Studies Depression Scale (CES-D)

The Center for Epidemiological Studies Depression Scale (CES-D) is a brief, unidimensional, self-report, depression scale developed to identify depression in the general population (Derogatis & DellaPietra, 1994; Gatz, Johansson, Pedersen, Berg, & C, 1993; McDowell & Newell, 1996; Mui, Burnette, & Chen, 2001; Radloff, 1977). It is a 20-item inventory to assess the respondent's perceived mood and level of functioning within the past 7 days. CES-D was developed in 1970s (Radloff, 1977). Items for CES-D were selected from previously validated depression scales, including e.g., BDI and MMPI. Respondents are asked to assess how often they experienced each symptom or positive phenomenon on a 4-point scale (0 - 3), all questions concerning the past week [7]. The scale can be found in (McDowell & Newell, 1996). Although CES-D is a self-report scale, it may be used as part of an interview. It was designed to cover major components identified in the literature:

---

[7] Example items of CES-D: 'I had trouble keeping my mind on what I was doing' (Psychomotor retardation), 'I felt depressed (Depressed mood)', 'I felt hopeful about the future' (Well-being), 'People were unfriendly' (Interpersonal difficulties).

depressed mood, feelings of guilt and worthlessness, feelings of helplessness and hopelessness, psychomotor retardation, loss of appetite, and sleep disorders. CES-D is nowadays widely used for screening depressive symptomatology among older populations (Gatz et al., 1993; Mui et al., 2001). CES-D is considered to be one of the best survey instruments for identifying depression.

CES-D does not assess severity of depression in a manner relevant to diagnosis or clinical assessment. When it is used as a screening instrument, a summary score above a certain cut-off point indicates a probable case of depressive disorder. The accuracy of the cut-off point has been discussed and it may need to be altered for use in different cultures (Mui et al., 2001). CES-D has been used effectively as a screening instrument in a number of community samples and its psychometric properties have been examined in numerous community-based studies. Furthermore, CES-D has become a standard indicator of depressive symtomatology among older adults (Mui et al., 2001).

Beck Depression Inventory (BDI)

The Beck Depression Inventory (BDI) is a unidimensional, self-report depression measure, and it has been in use over forty years. BDI was initially designed for use as a semi-structured interview tool for trained interviewers, but was later developed further to become a self-administered instrument (Beck et al. 1961,(Beck, Steer, & Garbin, 1988; Derogatis & DellaPietra, 1994; Katz, Katz, & Shaw, 1994; McDowell & Newell, 1996). The revised BDI was published in the 1970s. BDI is used both for psychiatric use (to assess current severity of depression) and for detection of depression in normal populations (Beck et al., 1988; Katz et al., 1994). Each of the 21 items consists of several statements evaluating the degree of given depressive symptoms and attitudes. 15 items cover emotions, four behavioral changes, and six somatic symptoms. Each statement in an item has a rating on a four-point scale from 0 to 3 in terms of intensity of that symptom (Derogatis & DellaPietra, 1994; Katz et al., 1994; McDowell & Newell, 1996). These scores are summed up to yield a total depression score. Beck postulates that the frequency of depressive symptoms is distributed along a continuum from non-depressed to severely depressed. There is also a short 13-item form of BDI (introduced in 1972).

There is a wide literature on validity and reliability of BDI (Beck et al., 1988; Katz et al., 1994). BDI has been reported to have high internal consistency and sufficient stability in terms of test-retest reliability for non-psychiatric patients. High concurrent validity has been shown in tens of studies by high correlations between BDI and well-established instruments, but the correlations have, however, been strongest with clinicians' ratings (Beck et al., 1988; Katz et al., 1994). However, content validity of BDI is not sufficient for making clinical diagnoses (according to DSM-III-R) of depression (Katz et al., 1994). There are guidelines for BDI cut-off scores with patients diagnosed as having 'depression' and its severity. Beck himself warns against rigidity in setting cut-off points: the cut-off score ranges depend on the nature of the sample and the purposes for which the instrument is being used; they should be chosen according to application (Beck et al., 1988; McDowell & Newell, 1996).

BDI is a copyrighted instrument (© by Aaron T. Beck 1987). Copies of the Beck inventory are available from The Psychological Corporation Texas, USA.

Self-Rating Depression Scale (SDS)

SDS, first published in 1965, is a 20-item, self-report, unidimensional measure of depression originally developed to quantify current depression in patients with depressive disorder (Derogatis & DellaPietra, 1994; McDowell & Newell, 1996). SDS has been used in clinical studies, as a screening instrument, and in cross-cultural studies. It comprises ten positive and ten negative items, all with a four-point scale from 'none or a little of the time' to 'most or all of the time'. The scale is a published one, that is, available to everyone free of charge (McDowell & Newell, 1996) and there are guidelines for interpreting results, i.e. cut-off points for depressive disorder and its severity. SDS has been widely used in geriatric populations and it has been shown to be valid in elderly people.

Hamilton Rating Scale for Depression (HRDS)

HRDS is a clinician rated tool addressing depression. It is a semi-structured clinical interview, which means that there are no specified questions, and that the interviewers should have good clinical experience (McDowell & Newell, 1996). An adequate interview takes at least 30 minutes and Hamilton recommended two independent raters who later compare scores. HRDS was first developed by Hamilton in 1960 and it contains 21 ratings each measuring a depressive symptom. A total score sums the responses from 0 to 52 with generally agreed cut-point of seven. The ratings can be found in (McDowell & Newell, 1996). Also the Hamilton Anxiety Scale (HAS) is available and has similar proceedings. Inter-rater reliability of HRDS is generally high and good standardized instructions for administration of the scale are available (A Structured Interview Guide for the HRDS, the SIGH-D). However, the reliability of HRDS across different samples or within a sample across time has found to be insufficient (McDowell & Newell, 1996). Furthermore, content validity of HRDS has been criticized, because it neglects cognitive and affective symptoms and favors somatic symptoms.

4.3.3 Measures of psychosomatic concepts

Various psychosomatic constructs can be measured by psychometric measures. These may be relevant in health survey settings to obtain information on the possible psychosomatic background of the physical complaints recorded in a survey. Below are some examples of concepts and their measures from comprehensive health surveys.

Somatization and hypochondriasis

Somatization and hypochondriasis may be defined as a persistent preoccupation with disease despite reassurance given after thorough medical examination (Fink et al., 1999). For instance the *Whiteley Index* is a measure addressed to these phenomena (Fink et al., 1999; Pilowsky, 1967). The original questionnaire from year 1967 includes 14 empirically derived items and a newer shortened version of a set of 7

items [8] (Fink et al., 1999). 'Somatization' as psychological distress that arises from perceptions of bodily dysfunction is also measured in some general mental health questionnaires, e.g., Symptom Check-List-90 (SCL-90) and Brief Symptom Inventory (BSI). In addition, the GHQ scales include 'hypochondriasis' as one of the four elements of distress.

Alexithymia

Alexithymia is a hypothetical personality construct associated with psychosomatic diseases, somatization disorders and some other medical and psychiatric illnesses (Bagby, Taylor, Parker, & Loiselle, 1990). Alexithymia is a multidimensional psychosomatic construct defined by the following cognitive-affective characteristics: a) a difficulty in identifying feelings and distinguishing them from the bodily sensations of emotion, b) a difficulty in describing feelings to others, and c) an externally oriented style of thinking (Gardos, Schniebolk, Mirin, Wolk, & Rosenthal, 1984; Parker, Bagby, Taylor, Endler, & Parker, 1993).

The Toronto Alexithymia Scale (TAS) was originally a 26-item self-report measure of alexithymia (Taylor et al., 1988). The three-factor model presented above has empirically been found from TAS (Joukamaa et al., 2001; Parker et al., 1993). An improved, revised version of the scale, the Twenty Item Toronto Alexithymia Scale (TAS-20), was introduced in 1994 (Bagby, Parker, & Taylor, 1994; Bagby, Taylor, & Parker, 1994). Its internal consistency, test-retest reliability, and convergent, discriminant and concurrent validity have been shown to be good (Bagby, Parker et al., 1994; Bagby, Taylor et al., 1994; Bagby et al., 1990; Parker et al., 1993). Items [9] of TAS-20 with their subdomain distribution may be found in (Bagby, Parker et al., 1994).

---

[8] Examples of items of *Whiteley Index*: 'Do you think there is something seriously wrong with your body?', 'Is it hard for you to believe the doctor when he tells you there is nothing to worry about?'.

[9] Example items of *TAS-20*: 'I am often confused about what emotion I am feeling' (Difficulty Identifying Feelings), 'It is difficult for me to find the right words for my feelings' (Difficulty Describing Feelings), 'I prefer talking to people about their daily activities rather than their feelings' (Eternal Oriented Thinking).

4.4   Quality of life measures

Quality of life (QOL) measures combine several well-being and health-related themes in one instrument and their aim is usually to cover physical, emotional, and social dimensions of health (McDowell & Newell, 1996). Both physical functioning and social and emotional aspects of health are evaluated. Many of these measures cover other areas in addition to these and have a very broad scope. The measures are usually and increasingly called 'quality of life' measures, although the term QOL is rarely and variously defined. These measures have also been called 'general health status measures' or with a more precise term 'health-related quality of life' measures. Quality of life issues measured by these instruments are health-related because they are usually used in medical or health settings or are based on health and functioning. Quality of life as it is used in clinical research is a vague term without conceptual clarity (McDowell & Newell, 1996) and often with an operational definition only. On the other hand, there are numerous measures addressing quality of life (McDowell & Newell, 1996) and, furthermore, quality of life measures are nowadays quite common in health surveys. QOL measures are in general thoroughly tested for reliability and validity. They represent successful application of test development procedures in health measurement (McDowell & Newell, 1996). Having an internationally validated quality of life assessment makes it possible to carry out quality of life research collaboratively in different cultural settings and to compare the results.

QOL does not describe mental health *per se*, but mental health (positive or negative) is definitely a part of quality of life. The two concepts are interrelated because mental disorders cause disruption in quality of life (WHO, 2001). Furthermore, QOL (when not defined as 'health-related' QOL) does not necessarily reflect a person's health status. Instead, aspects of well-being could be assessed separately from the health concept. Health is only one of the underlying domains of QOL. When assessed as a health related concept, QOL measures are often used for assessment of the impact of symptoms and disorders in life. In clinical medicine and health care, QOL measures are typically used as outcome measures.

QOL measures may be health profiles describing health status by a set of scores, or health indexes, summarizing health in a single number. The profiles evaluate different dimensions of quality of life, indicating that each domain should be

measured separately. The profiles have a clinical perspective and background and are often based on symptom checklists to monitor progress of an individual patient. Supporters of the health index approach demand that these several dimensions should be combined into an overall score to obtain an index with practical value.

McDowell & Newell (McDowell & Newell, 1996) reviewed twenty-one QOL measures, with their intended application areas and properties. Most of the scales were intended for clinical or research application, only 6 measures out of the 21 were meant to be used in survey settings. Furthermore, only three of these measures have been intended for survey use only, namely the two versions of the SF measure (SF-20 and SF-36) and one older time-consuming interviewer-rating measure (the Multilevel Assessment Instrument).

Three QOL measures, all of which have been widely used in national health surveys, will be briefly described below. These are the quality of life measure by WHO (WHOQOL), the SF Health Survey measures (SF-36, SF-20, SF-12), and the EuroQol instrument. The first two are health profiles and the last one a single index score.

The World Health Organization Quality of Life Assessment (WHOQOL)

WHOQOL is a health-related quality of life measure developed by the World Health Organization Quality of Life Group (The WHOQOL Group, 1995, 1998b; WHOQOL Group, 1994). Beyond the measure, there is the definition of quality of life as " *individuals' perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns*". It has been developed and piloted collaboratively in 15 culturally diverse centres, including four EU countries, in an attempt to develop a cross-culturally applicable measure.

WHOQOL produces a multi-dimensional profile of scores across six domains and 24 subdomains of quality of life (The WHOQOL Group, 1998b; WHOQOL Group, 1994). The domains are 1) physical domain (including pain and discomfort, energy and fatigue, sexual activity, sleep and rest, sensory functions), 2) psychological domain (positive feelings, thinking, learning, memory and concentration, self-esteem, body

image and appearance, negative feelings), 3) level of independence (mobility, activities of daily living, dependence on medication or treatments, dependence on non-medical substances, e.g., alcohol, tobacco, drugs . . . , communication capacity, working capacity), 4) social relationships (personal relationships, practical social support, activities as provider/supporter), 5) environment (physical safety and security, home environment, work satisfaction, financial resources, health and social care, opportunities for acquiring new information and skills, recreation/leisure activities, physical environment, transport), 6) spirituality/religion/personal beliefs. Within each domain several sub-domains of quality of life summarize that particular domain. Individual questions of WHOQOL may be found in (The WHOQOL Group, 1998b).

WHOQOL-BREF is an abbreviated version of the WHOQOL-100 (The WHOQOL Group, 1998a) and includes 24 items. It includes four main domains: physical health, psychological, social relationships, and environment. WHOQOL-BREF has been reported to possess the same psychometric properties as the longer WHOQOL-100. The development of WHOQOL has three important features. First it involves a collaborative development approach to international instrument development. With this approach standardization, equivalence between settings and translation issues were at the forefront of the development process with genuine international collaboration. Second, in the development of WHOQOL there was iterative input of quality of life researchers and revision of it at the grassroots level of the development process. The third important feature was the use of the tried and tested WHO translation method (The WHOQOL Group, 1995).

The Short-Form Health Survey (SF-36, SF-20, SF-12)

The Short-Form Health Survey Index of Quality of Life is a development from the large American Medical Outcomes Study (MOS) (Bullinger et al., 1998; Gandek et al., 1998; Gandek et al., 1998; Ware et al., 1998). It is derived from the work of the Rand Corporation of Santa Monica USA during the 1970s. It is a multi-purpose health survey and a generic measure of health status, i.e., not addressed to a specific group of people (McDowell & Newell, 1996; Ware et al., 1998). SF measures were developed for use in population surveys and evaluative studies of health policy; they

were developed to measure population health. The SF measures may be self-administered or used in personal or telephone interviews.

There are several versions of SF measures, and the number of their items is indicated in the name of the measure. Nowadays, the most frequently used version is SF-36. It was designed in response to criticism that the previous 20-item version was too limited in scope. SF-20 was published in 1988, eighteen of its items were taken from the previous Rand scales. Thus SF-20 is a predecessor of SF-36, it is not a part of SF-36, which should be noted when evaluating comparability between the two. Because large general population surveys may only have room for a one page questionnaire, SF-12 was developed to reproduce the SF-36 physical and mental health summary measures with fewer items. Thus SF-12 is a part of SF-36 (Gandek et al., 1998). Currently, a short SF scale for survey purposes, SF-8, is in development [10]. SF-8 covers the same dimensions as SF-36 and SF-12. By The International Quality of Life Assessment (IQOLA) Project, SF measures have been widely and cross-culturally validated (Bullinger et al., 1998; Gandek et al., 1998; "Journal of Clinical Epidemiology, Special Issue on SF-36 measure," 1998; Sullivan et al., 1995; Ware et al., 1998), and meant to be standard measures.

SF-36 covers eight main health domains as well as the summary measures of physical and mental health. The eight domains represent the concepts most frequently measured in health surveys and affected by disease and treatment. These domains are divided into four physical health scales (Physical Functioning, Role-Physical, Bodily Pain, and General Health) and four mental health scales (Vitality, Social Functioning, Role-Emotional, and Mental Health). SF-12 is divided only into physical and mental domains. In the SF-8 scale under development, one item represents each of the eight scales. For copyright reasons, exactly the same items as in SF-36 have in some countries been published and validated as a measure called RAND-36 (Aalto, Aro, & Teperi, 1999). SF-36 and SF-20 may be found in (McDowell & Newell, 1996).

---

[10] The SF-8 was constructed to replace the SF-36 and SF-12 in population health surveys in the U.S. and internationally. Accordingly, it has been translated and linguistically validated for use in more than 30 countries and languages using IQOLA Project methods. The development, validation and norming of the new SF-8 Health Survey is documented in a manual by John E. Ware, Mark Kosinski, James E. Dewey and Barbara Gandeck: "How to Score and Interpret Single-Item Health Status Measures: A Manual for User of the SF-8 Health Survey", Quality Metric, Inc., Lincoln RI, USA (in press).

The EuroQol quality of life scale (EuroQol)

The EuroQol scale measures generic health status as a single index score (McDowell & Newell, 1996). EuroQol has been developed for evaluative studies and policy research, not for national health surveys. The European Quality of Life Group, an international collaborative program with members from five countries, developed the measure in the late 1980s. They defined a core set of generic health related quality of life domains. It was intended that these items may be supplemented in some studies, but that the core set would permit cross-national comparisons and encourage methodological standardization. As a result, EuroQol includes only core functional status questions and no questions on symptoms. It covers five dimensions of health: mobility, self-care, role (or main) activity, pain, and mood (an earlier version included six dimensions). Within each dimension, the respondent chooses one of three alternatives, indicating 'no problems', 'some problems', and 'severe problems' in this dimension [11]. Contrary to the health profile measures presented above, EuroQol is a health index measure. In practice, this means that there are only five questions to be answered and thus the EuroQol index is quite time-saving measure. However, its main disadvantage is that there are only two possible severity levels whereas most other measures use three.

EuroQol was developed simultaneously in Dutch, English, Finnish, Nowegian, and Swedish. Translations into other European languages are available. The scale may be found in (McDowell & Newell, 1996).

---

[11] In *EuroQol* dimension Self-care the alternatives are 'I have no problems with self-care', 'I have some problems with self-care', 'I am unable to wash or dress myself'.

4.5 Other measures

4.5.1 Occupational health measures

The Maslach Burnout Inventory (MBI and MBI–GS)

MBI has become an almost universally accepted gold standard to assess burnout (Schutte, Toppinen, Kalimo, & Schaufeli, 2001). 'Burnout' is defined by the test authors as a syndrome of emotional exhaustion, depersonalisation, and reduced personal accomplishment. MBI measures these three components of burnout: 1. Emotional exhaustion: no longer being able to give oneself at a psychological level, 2. Depersonalisation: negative, cynical attitudes and feelings about one's clients, and 3. Reduced personal accomplishment: a tendency to evaluate oneself negatively, particularly with regard to one's work with clients (Prosser et al., 1999). Burnout is suggested by higher 'emotional exhaustion' and 'depersonalisation' scores, and by lower 'personal accomplishment' scores. Emotional exhaustion is the key aspect in the syndrome and refers to feelings of being over-extended and drained from one's emotional resources (Schutte et al., 2001).

The original MBI has been designed for use in the human services, i.e., persons working with people. To measure burnout in occupational groups outside the human services an extra version was developed, the Maslach Burnout Inventory – General Survey (MBI–GS). MBI–GS was developed simultaneously in The Netherlands, Canada, and Finland. This scale has been recently validated (Schutte et al., 2001). MBI–GS comprises three distinctive, yet related, subscales (parallel to those of the original MBI): Exhaustion (five items), Cynicism (five items), and Professional Efficacy (six items). All items [12] are scored on a 7-point scale from 0 (never) to 6 (daily).

---

[12] Example items of *MBI-GS*: 'I feel used up at the end of the workday' (Exhaustion), 'I have become less enthusiastic about my work' (Cynicism), 'In my opinion, I am good at my job' (Professional Efficacy).

4.5.2 Measurement of social health

Social health is less familiar and less frequently assessed than the topics of physical and psychological health (McDowell & Newell, 1996) although it is one part of the entity of health with physical, psychological, and social dimensions (WHO, 2001). Social health may be regarded as comprising characteristics of society or it may concern measurements of individuals (McDowell & Newell, 1996). It deals with social and cultural environment: social networks, social support, social isolation, and life events.

Social health has been defined to be "that dimension of an individual's well-being concerning how he or she gets along with other people, how other people react to him or her, and how he or she interacts with social institutions and societal more"(McDowell & Newell, 1996). This definition includes elements of personality and social skills, and it also reflects the norms of the society in which the individual lives. Most measurements of social health do not deal with 'health', but instead 'well-being', 'adjustment', 'performance', or 'social functioning'. In fact, human interaction is evaluated.

After the WHO three-dimension definition of health emerged, there has been an emphasis on treating patients also as social beings who live in a complex social context (McDowell & Newell, 1996). It was discovered that those who are well integrated into their communities tend to live longer and have an increased capacity to recover from disease. On the other hand, social isolation is a risk factor for disease. The social view of medicine postulates that the aim of care should be to reintegrate patients into productive lives rather than to treat their symptoms only. This has resulted in a demand for new instruments to assess need for care and outcomes of care. There are also practical reasons for measuring individual's social well-being and adjustment, for instance the high costs of institutional care, resulting in a need to evaluate the patients' abilities to live in the community. This is especially important in regard of the mental health. Measurements on these issues may be applied in health care including outcome of rehabilitation: has the patient returned to a productive position in the society.

Another reason to evaluate social health is to investigate the influence of social support and social integration on an individual's physical and psychological well-being (House, Landis, & Umberson, 1988). In these cases social health factors are not outcome variables but determinants of health.

Social health may be instrumentalized, e.g., as *social adjustment*, *social roles*, *social networks*, and *social support* (for some important measures of the area, see (McDowell & Newell, 1996)

The concept of *social adjustment* in the health field derives from psychiatry (McDowell & Newell, 1996). Problems in personal or social relationships and communication are the most common reasons for seeking mental health care. A person's social adjustment may indicate need for care and also be an indicator of outcome. Use of adjustment scales in psychiatry reflects a shift from concepts of disease and deviance towards a person's capabilities in social integration and adjustment. Social adjustment has been broadly defined as the interplay between the individual and his social environment and his success in his *social roles*. Social adjustment may be measured by investigating these two areas: as a person's satisfaction with his relationships or as his performance in various social roles. Thus affective responses, such as unhappiness and anxiety are measured. Measurements of social adjustment and evaluations of concepts like, e.g., life satisfaction or quality of life are not clearly separate. Therefore, the scales in this area are for practical reasons often put under the heading of "subjective well-being". However, social adjustment scales clearly contain evaluation of the well-being related to social relationships.

Social support is an interactive process in which particular actions and behaviors may have a positive effect on an individual's social, psychological, and physical well-being (Cohen & Wills, 1985; O'Reilly, 1988). Social support involves emotional concern, instrumental aid, information, and appraisal (Dunkel-Schetter, Folkman, & Lazarus, 1987). The person's own subjective perceptions of the support received are positively related to psychological distress and physical health outcomes (Cohen & Wills, 1985). Social support appears to be one of the mediating factors in the relationship between stress and health outcomes, and social support including interventions for stressed individuals seem to help in diminishing the consequences

of stress or to improve adjustment (Dunkel-Schetter et al., 1987). *Social network* refers to the quantity of social relationships forming the sources of social support. Both social networks *per se* and social support derived from them have been shown to be predictors of physical health (Cohen & Wills, 1985).

PART II

5 MENTAL HEALTH MEASUREMENT IN NATIONAL HEALTH INTERVIEW (HIS) AND HEALTH EXAMINATION (HES) SURVEYS IN EU/EFTA COUNTRIES

5.1 Health interview and health examination surveys in EU/EFTA countries

In the first phase of the EU Health Monitoring Program (HMP) the project 'Health Surveys in the EU: HIS and HIS/HES evaluations and models', collected information on the methods of the health interview (HIS) and health examination (HES) surveys which was read into the health survey database (HIS/HES database, 2000). The database has since been developed further (HIS/HES database, 2001, 2002; The HISHES Group, 2003) and is now available via internet (The HISHES Group, 2003).

The inventory of national health interview surveys in the EU, with Norway, Iceland and Switzerland, showed that HISs are being executed regularly in most countries (HIS/HES database, 2001; Hupkens & Swinkels, 2001). Until 2001, HIS had been carried out in 14 of the 18 EU/EFTA Member States (Hupkens & Swinkels, 2001). In three of the four countries without regular HIS, multi-purpose surveys, typically censuses or level of living surveys, covering health and/or surveys on a specific health theme were conducted.

Health interview surveys and questionnaires are a good way to obtain data on e.g. symptoms experienced by the subjects or their health related behaviour. Clinical measurement is, however, often needed to obtain more information about many chronic conditions, such as mental health problems. Such data can only be obtained by carrying out health examination surveys or by supplementing health interviews by health examination methods. The division into health interview and health examination surveys is of course not clear cut, in practice all HESs include a HIS component.

In contrast to HISs, national health examination surveys are relatively rare in Europe. Population based, national, rather comprehensive HESs have, until 2001, been conducted only in five EU countries: Finland, Germany, Ireland, Netherlands and UK (Koponen & Aromaa, 2001, Table 1). In Spain (Catalonia), a large regional HES has

been conducted in 1996. The survey samples vary considerably. In addition to national comprehensive surveys, several local, regional and focused surveys have been carried out. In fact, target-focused and geographically limited HES surveys have been conducted in almost all EU/EFTA countries. Quite often they have concerned cardiovascular disease risk factors such as the MONICA risk factor surveys. One target for focused HES surveys has been the evaluation of mental health. Also international collaborative mental health surveys have been conducted in EU/EFTA Member States.

In this report, we will review only mental health and quality of life components of **the national comprehensive HESs and HISs recorded in the HIS/HES database**. Additional measurements on, e.g., cognitive functioning, substance abuse, and eating disorders will not be considered.

5.2 Mental health measurement in national health surveys

5.2.1 The surveys and source of information

The initial search concerning the use of mental health measures in national health surveys was performed from the first version of the database (HIS/HES database, 2000). The examination suggested that a variety of methods (validated measures, adapted versions of validated measures, methods developed for purposes of the current survey, single questions) have been employed. If validated methods have been used, they may have been adapted for use in a single survey only. In these cases, various aspects of validity of the measurement instrument have often not been assessed. On the other hand, some surveys have employed internationally validated measures, i.e., the validation processes have been carried out at international and population level, providing a strong basis for international comparison. The heterogeneity of the use of mental health measurement, the varied targets of the measurement, and the variety of the instruments results in poor between-study and international comparability.

The first version (year 2000) of the database did not provide sufficient information on the mental health instruments used in the surveys. In the year 2001 version

(HIS/HES database, 2001) [13], there were 47 surveys, out of which eight included a HES component. The HIS section of the database contains a list of 'Standardized or recommended instruments', including information on the use of GHQ (a general mental health measure) and one of the SF measures: SF-36/20/12 (a quality of life measure). There is also an open section for 'Other measures' used. According to the 2001 version, GHQ was used in the original form in seven surveys and in an adapted form in two more surveys. SF measures had been used in four + one surveys. The section Other measures included a variety of measures on various issues, also measures on mental health and quality of life.

The search was repeated from the newest, the website version of the database (The HISHES Group, 2003). We focused on the surveys conducted in years 1998 – 2002. In the database, both individual mental health and quality of life questions (section Questions) and validated instruments (section Methodologies) are covered. The questions are grouped into the domains Quality of life, Aspects of mental health, General mental health, Positive mental health (self-esteem, mastery, coherence, self-efficacy) and Social health, while also the domain Self-assessed perceived health includes questions describing these issues. The instruments are grouped as in the earlier version of the database (see paragraph above). At the moment the database does not include information on the version (number of items) of the measure or the specific subscales used when the whole instrument has not been applied. The search was completed *via* some personal communication with contact persons of the surveys. Concerning these parts the database is still under development.

5.2.2 Mental health instruments used

The results showed that measurement of mental health and quality of life with validated instruments is not always included in comprehensive national health surveys (Table 5.2.1). Furthermore, the area is very heterogeneously covered and when mental health measurement is present, its contents vary, i.e., the targets of measurement differ. Also the number of mental health measurement targets in a

---

[13] From this version we adopted relevant surveys, the mental health and quality of life instruments used in these surveys, and contact persons of the surveys, and completed the search by mailing a questionnaire (Appendix 1.). These results are going to be used for improving the database.

survey vary considerably. The use of mental health and quality of life instruments in years 1998 – 2002 in the surveys by country is presented in Table 5.2.1.

In the HIS/HES database, there were a total of 64 surveys from 18 European counties. Eleven of the surveys had had also a HES component (six were from the same country: United Kingdom) and one was a plain HES. Twenty-one surveys (33%) included some mental health or QOL measurement with validated instruments. The percentage was the same in the year 2001 search. The remaining two-thirds of the surveys had either neglected the area, had evaluated it by individual questions or had applied validated measurement in subsamples only. According to the database, 12 European countries had applied mental health or QOL instruments in their surveys. The number of these measures in a survey varied between one nad eight, most of the surveys used only one such measure (usually either a general mental health scale or a health-related quality of life scale). Some surveys included a large number of measures, namely Health 2000 from Finland and Survey of Lifestyle, Attitudes and Nutrition (SLÁN) from Ireland.

Most often, the surveys measured general mental health and the tool for this was the GHQ-12 scale. In some surveys, the SCL-90 was used. Health-related quality of life was evaluated as often as general mental health, the most common measure was one of the SF questionnaires, occasionally also the WHOQOL or the EuroQol. The General Health Questionnaire (GHQ-12) was the most frequently used mental health measure (fr. 12). The most frequently used quality of life measure was the SF-36 (fr. 9).

Table 5.2.1 Mental health and quality of life instruments in comprehensive national HIS and HES surveys [1] in EU/EFTA countries, by country .

Country	Survey	Year	Instruments yes/no	Name of instrument original form/ adapted
Austria	Microcensus	1999	no	
Belgium	Health Interview Survey	2001	yes	GHQ orig.
				SF 36/20/12 adapt.
				SCL-90
	General Socio-Economic Survey	2001	no	
Denmark	Health and Morbidity in Denmark	2000	yes	WHOQOL
				SF 36/20/12 orig.
				EuroQol
Finland	Survey on health behaviour	2000	no	
	Health 2000 (HIS/HES)	2000	yes	GHQ orig.
				EuroQol
				M-CIDI adapt.
				BDI adapt.
				SLC-90 (somatization) orig.
				MBI-GS orig
				TAS-20 orig.
				Whiteley Index
	Health Behaviour Survey among the Adult Population	2001	no	
	The European Community Household Panel Survey	2001	no	
	Health Behaviour Survey among the elderly population	2001	no	
	The National Finrisk Study (HIS/HES)	2002	no,	in psychosocial subsample, 60% of the total, yes
France	Health and Social Protection Survey	1998	no	
	Handicaps, Disabilities and Dependency Survey	1999	no	
	Health Barometer	1999	yes	GHQ orig.
	French survey on living conditions and aspirations	1999	no	
	Continuous survey on households living conditions	2000	no	
	Survey on households living conditions	2001	no	
	The INSEE Survey on Handicaps, Disabilities and Dependency	2001	no	

---

[1] (The HISHES Group, 2003)

	French survey on living conditions and aspirations	2001	no	
	Health and Care Interview Survey	2002	no	
	National survey on health and national health insurance (ESPS)	2002	no	
Germany	Survey on living conditions, health and environment	1998	no	
	German National Health Examination and Interview Survey (HIS/HES)	1998	yes	SF 36/20/12 orig. M-CIDI adapt.
	Questions on Health	1999	no	
Greece	National Greek Survey: Psychosocial factors and Health	1998	yes	SCL-90 adapt. CES-D adapt.
Iceland	Health and lifestyle of the Icelandic population	2001	yes	GHQ orig.
Ireland	Survey of Lifestyle, Attitudes and Nutrition (SLÁN) (HIS/HES)	1998	yes	EuroQol orig.
	Living in Ireland Survey	2000	yes	GHQ adapt.
	Living in Ireland Survey, 2001	2001	yes	GHQ adapt.
	Survey of Lifestyle, Attitudes and Nutrition (SLAN) (HIS/HES)	2002	yes	EuroQol orig. GHQ-12 orig. WHOQOL –BREF orig. CES-D orig.
Italy	Health Conditions and the Use of Health Services	1999	yes	SF12 adapt.
	Aspects of daily living	2000	no	
	Aspects of daily living	2001	no	
Luxembourg	Eurobarometer	1999	no	
	European Community Household Panel	2000	no	
	European Community Household Panel	2001	no	
	Socio-Economic Panel Living in Luxembourg	2001	no	
	2002 LFS ad hoc module employment of disabled people	2002	no	
	Eurobarometer October 2002	2002	no	
Netherlands	Continuous Quality of Life Survey	1998	yes	Affect Balance Scale adapt.
	Patient survey - Second Dutch National Survey of General Practice	2001	yes	GHQ orig. SF 36/20/12 orig.
	Continuous Quality of Life Survey	2001	yes	SF 36/20/12 orig.
	Netherlands Health Examination Survey, Regenboog (HES)	2001	no	

Norway	Survey on Living Conditions	1998	yes	SF 36/20/12 orig.
	Survey on living conditions & health, care and social relations	2002	yes	SF 36/20/12 orig.
Portugal	National Health Interview Survey	1999	no	
	General Census 2001	2001	no	
Spain	Impairments, Disabilities and Health Status Survey	1999	no	
	National Health Survey	2001	no	
Sweden	Living Conditions Survey	1999	no	
	Living Conditions Survey	2001	no	
Switzerland	WHO - Multi-country Survey Study on Health & Responsiveness	2001	no	
	Swiss Health Survey 2002	2002	no	
	World Health Surveys-Consolidated multi-country questionnaire	2002	no	
United Kingdom	The Scottish Health Survey (HIS/HES)	1998	yes	GHQ orig.
	Health Education Monitoring Survey	1998	no	
	Health Survey for England (HIS/HES)	1998	yes	GHQ orig.
	Health Survey for England (HIS/HES)	1999	? [2]	
	General Household Survey	2000	no	
	The Health Survey for England (HIS/HES)	2000	yes	GHQ orig. SF 36/20/12 orig.
	Census 2001	2001	no	
	The General Household Survey	2001	no	
	Health Survey for England (HIS/HES)	2001	yes	GHQ orig.
	Health Survey for England (HIS/HES)	2002	?	
	General Household Survey	2002	no	

---

[2] Information not yet available in the database.

5.2.3 Mental health measures in the health examination surveys

In the HESs, the use of mental health and quality of life measures was much more common than in HISs (Table 5.2.1). Most of the surveys with health examination included some measurement of the area. Several kinds of measures were used: diagnostic measures, general mental health questionnaires, depression scales, quality of life instruments, etc. Application of these measures was, however, heterogeneous: some surveys measured mental health and well-being widely, while others neglected the area. Two surveys used a diagnostic mental health package (M-CIDI). Psychological measures for psychiatric morbidity - General mental health questionnaires and Depression scales - were commonly applied. Also health-related quality of life instruments were quite often included in the survey measurement. Sometimes also other measures were applied, targeting e.g., psychosomatic complaints, occupational health, and social health.

Health 2000 included a larger number of mental health related measures than any of the other surveys. In the Irish Survey of Lifestyle, Attitudes and Nutrition (SLÁN), there also were several mental health and quality of life measures. In addition, one HES, The National Finrisk Study, included a large psychosocial subsample, covering as much as 60% of the respondents.

In practice, all HESs comprised some validated mental health or quality of life measurement. Although these evaluations were common, only few measures were used in more than one HES, namely the GHQ-12, the EuroQol scale, and the diagnostic M-CIDI package. The GHQ-12 questionnaire was used in HES surveys of three countries (Table 5.2.1) and, furthermore, repeated several times in the U.K.. CIDI and EuroQol were used in two countries. In addition, WHOQOL was used as a whole in one country, and partly and in a subsample of another. The GHQ-12 was used in the Finnish Health 2000, in the Irish Survey, and in the repeated UK surveys. In the Finnish Health 2000 and in the German HES, the same computerized version of M-CIDI was applied. EuroQol and WHOQOL were used in Health 2000 in Finland and in the Irish HES (in Health 2000 the WHOQOL short form only in persons aged 18-29 years). SF-36 was applied in Germany and repeatedly in U.K.

Health-related quality of life measurement was often included in HESs. The instruments varied: EuroQol, WHOQOL and SF-36 (used also in many HIS surveys) were all applied for evaluating the concept of health-related QOL.

5.3 Targets of mental health measurement in national health surveys

In the investigated five-year period, mental health issues were evaluated with validated instruments only in the minority of the comprehensive European national health surveys. There is no general agreement on that the psychological, social, or mental health dimensions of health should always be included in a comprehensive health survey. However, if mental health had been surveyed, there seemed to be some agreement on the aspects – and even on some measures – to be included. General mental health and health-related quality of life were the most commonly evaluated mental health domains, and GHQ-12 and the SF measures the most commonly applied mental health instruments. The other targets of measurement were heterogeneous and covered areas from depression and some other specific psychiatric disorders, to personality traits and aspects of social health.

Mental health was much better covered in health examination than in health interview surveys. This was not, however, due to the examination approach itself, because also self-report questionnaires of the area (applied sometimes by postal survey) were much more common in HESs than in HISs. Generally in HES settings there are more resources and more expertise than in HIS. Two national HESs used a large psychiatric diagnostic package (CIDI) currently applicable only in HES settings. The CIDI-SF offers a promising alternative for HIS use.

Major public health concerns, and particularly depression, were usually included if a mental health component was present. Depression (affective disorders) was measured, with varied methods, in most of the HISs that used mental health measures and in all such HESs. Evaluation of depression is typically a subscale in the general mental health questionnaires. Affective disorders have also earlier been reported to be the most commonly assessed mental disorders in health surveys. There is a strong tradition in measuring depressive symptoms and clinical depression, and specific depression scales are easily available and they are probably

the best validated mental health measures. Measurement of psychological and social aspects of health has usually not been included in national European health surveys.

Out of the surveys reviewed, 21 of 64 surveys comprised validated mental health measurement. If such evaluation exists, the measures were usually limited to short scales of general mental health and/or health-related quality of life. Psychological morbidity is the measurement target in the 'general mental health' instruments when psychological well-being is targeted in the quality of life measurement. There was a tendency to include in a survey a good validated measure of the area. In HISs, this was either a general mental health measures and/or a health-related quality of life questionnaire, and in HESs, in addition, also diagnostic measures for psychiatric morbidity, a separate depression inventory, and/or additional quality of life scales, completed occasionally with various psychological and social health scales. If mental health was part of a survey, depression was almost always one of the targets.

5.4 Comparability of the measurement across surveys

Several organizations carry out national health surveys and apply standardized measurement in order to enhance international comparability (ECHI Project, 2001; Hupkens & Swinkels, 2001; Koponen & Aromaa, 2001; WHO, 1999). Many international projects, notably those of the EU health monitoring programme and WHO, have attempted to improve standardization of mental health measurements. Some of these initiatives have produced somewhat contradictory recommendations. International comparability requires 1. use of valid instruments, which are 2. applicable in survey settings in different countries, 3. and that these instruments have been used in a similar way.

It was unfortunate to discover that in the current European health surveys the validated measurement of mental health was rare, the targets of measurement varied, and the instruments used differed. Thus, there are few data for international comparison from comprehensive national surveys. It should be noted that the mental health specific surveys that are conducted in certain countries with strong expertise and tradition most likely yield comparable results. These have been carried out separately from the comprehensive national surveys. However, also the national

comprehensive surveys provide some opportunities for international comparison as described in the next paragraph.

5.4.1 Comparable instruments

Two often applied measures: the GHQ (usually the shortest version, the GHQ-12) and the SF measures (usually SF-36) enable direct international comparison of mental health aspects across several European countries. Also the use of WHOQOL and EuroQol quality of life scales offer some opportunities for comparison. The few studies employing the CIDI package allow good possibilities for comparison of the prevalence of mental health diagnoses across countries.

SF-36/20/12/(SF-8)

The SF measures are internationally validated instruments, although their origin is not in survey work. SF-36 has been widely used in health surveys, but may be too extensive to be used in comprehensive health surveys and therefore SF-12 was developed to overcome this problem. However, there seems to be a tendency to use only some of the eight parts of the SF-36 in survey use. The recent HIS recommendations of the area have recommend the use of certain subscale or subscales only, targeting psychological distress, psychological well-being, and impairment (Lehtinen et al., 2001). The development effort of the SF measures has been ambitious because the intention was that they become gold standard instruments for the measurement of health-related quality of life. The SF measures continue to be developed, and at the moment the aim is to develop and validate a short SF-8 version for survey use.

WHOQOL and EuroQol

The quality of life scales of WHO and the European Quality of Life Group are internationally validated scales, but neither of them is developed for health survey use. The original WHOQOL with 100 items is not applicable for comprehensive surveys, while the WHOQOL-BREF, the abbreviated version with 24 items, might be usable. Contrary to the health profile measures (The SF measures and WHOQOL), EuroQol includes only five questions. This is useful for survey use. The unadapted

questions of EuroQol offer opportunity to compare components of QOL between European surveys.

GHQ-12

GHQ-12 is no doubt capable of giving important information about the four major elements of current distress: depression, anxiety, social impairment, and hypochondriasis for survey purposes. It's aim is to identify persons with mental distress. Although not an internationally developed instrument, GHQ-12 has shown high validity when administered in several languages and countries. GHQ-12 is short and simple to administer. It is a good alternative, when a short and valid general mental health measure is sought with possibilities for international comparison of the results.

CES-D

The CES-D depression questionnaire has been developed originally for identifying depression in the general population. With 20 items it assesses the respondent's perceived mood and also the level of functioning. The items of CES-D have a longer history than the measure *per se*, they were selected from previously validated depression scales, e.g., BDI and MMPI. CES-D has not been internationally developed but its psychometric properties have been widely examined. CES-D has become a standard measure of depression among older adults. It is considered to be one of the best survey instruments for identifying depression.

CIDI

The use of the diagnostic CIDI package is giving promise for international comparison in the field of psychiatric morbidity. However, use of the CIDI main package is time-consuming and applicable only in surveys with an examination component and a personal interview. On the other hand, the short form of CIDI (CIDI-SF) is applicable also in HIS settings. It has recently been recommended for use, at least in part, in the interview surveys, targeting major depression and generalised anxiety disorder (GAD) (Lehtinen et al., 2001). The CIDI package was used in a computer assisted version in two European comprehensive HESs. The method is promising and can be applied in different populations.

Other measures

Several instruments have been used in the surveys to evaluate various concepts from the fields of psychiatric disorders, personality traits, or aspects of social health. Some of these instruments have been carefully validated and widely used in psychological and social research and are thus highly valid. Unfortunately, comparisons between surveys and countries are not possible, because they have been rarely used in several surveys.

5.4.2 Comparable measurement targets

The surveys seem to have been interested in major public health concerns and health-related quality of life. Furthermore, the surveys attempt to measure the same aspects of mental health and QOL. Parts of the concept of quality of life are well covered and comparable when the same instrument, usually SF-36, or the same subscales of the instrument have been applied. Addressing general mental health, GHQ includes the important concepts of depression, anxiety, social impairment, and hypochondriasis. These are genuinely comparable between the surveys, although they are only narrowly covered with the short methods. Depression and anxiety have been measured by using subscales in GHQ, SCL-90 and CIDI, and depression also by using two specific depression questionnaires, BDI and CES-D. Thus depression was commonly included in the surveys, but it was altogether measured in health surveys by five different methods, either specific instruments or subscales of instruments. However, these depression scales are partly related: items for CES-D have originally been taken from previous scales, including BDI. It may also be possible to evaluate trends in the concepts measured, although the methods have not been exactly the same. The prevalence of depressive disorders is likely to be roughly comparable.

5.4.3 Adapted measures

In many surveys, the instruments were adapted for the purposes of that investigation (The HISHES Group, 2003). It depends on the modification whether the adapted scale or instrument continues to be valid. Even tiny modifications may alter the validity of the whole measure. Most likely the validity of a subscale containing a modified item has also been modified. If single items have not been changed although only some have been included, results of these can be compared although the summary measure is lost. However, using single items from a validated measure is rarely useful since validity has been evaluated for the whole measure (or submeasure) and not for individual questions. Adding extra items is not a problem.

5.4.4 Summary

There are two main possibilities to improve the international comparability of mental health measurement. One is to create a comprehensive international recommendation on these issues, i.e., to standardize the measurement of mental health in health survey settings. The other is to start from the current use of instruments allowing comparison, enhance their use in present and future surveys, and hereby collect internationally comparable data. The most appropriate alternative is probably to use both approaches simultaneously. In both cases, application of the methods calls for international collaboration of experts.

International comparison of mental health measurement is possible if the instruments are the same, and, if they have been validated in each of the countries or (sub)cultures in which they are used. Therefore instruments should either be internationally developed or they should be later validated in different countries. It is also important to survey mental health in many more countries than today.

5.5 Conclusions related to the empirical information

Mental health measurement was rare in comprehensive health surveys. The substance of mental health measurement varied, i.e., there were varying targets of measurement. Also, the same targets were investigated by different methods. The number of targets and methods in a survey varied considerably.

The findings suggest that:

1) the use of validated mental health and quality of life measures is rare in comprehensive national HIS and HES surveys in Europe.

2) Targets of the measurement usually cover general mental health and health-related quality of life, and occasionally varying other targets.

3) Despite the application of two widely used instruments, the use of the measures is heterogeneous: variety of the measures and/or versions of the measures is broad.

4) Instruments have commonly been adapted for local use.

5) As a result, international comparability of the findings is limited to some instruments only.

6) There is a considerable need to improve the comparability of measurement in the field of mental health and quality of life in national HIS and HES surveys in Europe.

In order to improve validity and comparability several approaches are necessary. First, it is important to build on the survey instruments used and known to yield comparable results. Second, a set of recommendations and guidelines needs to be developed concerning mental health and quality of life measurement in surveys. Third, a network of experts should promote the choice and use of suitable measures in surveys. In addition, it would also be useful to carry out international comparisons on the basis of existing survey data. The implementation of a Europe-wide survey or a set of national surveys with sufficient standardization should be encouraged.

5.6 Recommendations on mental health and quality of life measurement in national HISs and HESs

Most people working with health surveys would probably appreciate a guide-book of recommendations concerning the "best buy" mental health measures. Such judgement, however, requires knowledge of details about the planned survey. The measures must in any case be assessed against the main interests of the survey, and only after that can the most suitable measures for the current purpose be chosen.

Some good advice for evaluating and choosing methods in health research may be found in (McDowell & Newell, 1996). Their checklist indicates that the following should be considered. 1) The purpose of the method is fully explained and the method found to be appropriate for the intended use. 2) The method is broad enough for the application (neither too long nor short) and preferably capable of identifying also positive health. 3) The conceptual approach has been clarified and is relevant to the measurement topic. 4) The method is feasible to administrate (how long does it take, is it self-administered/ is professional expertise required, does it use readily available data, is it acceptable to respondents). 5) The way of scoring is clear and it is suited to the statistical analyses planned. 6) The degree of change detected is adequate for the purpose, e.g., does it provide quantitative data. 7) The evidence for reliability and validity is strong enough.

If cross-survey or international comparability is aimed at, it is not enough that the measurement method is suitable for use in the survey in question. The same principles in selecting instruments should be considered in all the surveys meant to give information also in international settings. Comparability of results enhances their usefulness. As pointed out before (McDowell & Newell, 1996) there is, also in national surveys in Europe, considerable variation in the quality, sophistication and applications of the mental health measures and measurements. In particular, comparability of findings leaves much to be desired. However, health measurement has benefited from the theoretical and technical advances in test (or measure) construction already achieved in the social sciences (McDowell & Newell, 1996). The application of this knowledge to health measurements requires time and resources.

Recently, an important step in evaluating and comparing instruments has taken place (McDowell & Newell, 1996). There are now psychometric data available for most scales, i.e. at least some aspects of validity and reliability of the measures have been documented. There also seems to be a tendency towards using internationally developed and validated measures. Recently growing attention has been given to scoring health measures: more and more measures use formal scaling methods. The conceptual definition of the topics has improved. Nowadays, there may even be an effort towards "a science of health measurement" equivalent to psychometrics and econometrics (McDowell & Newell, 1996). *The methodology of the area should be known, and the knowledge applied to the research work, for research resources are wasted if studies do not use the best available techniques, and the accumulation of scientific evidence is delayed if comparable measurement methods are not used.*

Guidelines for developing psychosocial health measurements and measures should – also in survey settings – follow the same general rules and standards applied in psychological testing. They can be found in American Psychological Association (APA) 's publications (APA, 1985, 1999; McDowell & Newell, 1996). In summary, the standards state that

1) An article or manual should fully describe the method,
2) The definite version of the method must be readily available for users, preferably at no extra cost,
3) The name of the measure should describe its contents,
4) The conceptual definition behind the measurement should be given,
5) The origin of the questions should be given,
6) If there are revisions to the method, they should be clearly explained and psychometric data on the latest version presented,
7) Instructions should ensure standard administration,
8) Every measure should include reference scores from several populations (ideally a population sample of healthy people and samples from various patient groups),
9) The validity and reliability testing should include both the structure of the method and its relation to alternative measurements of the concept,
10) The measure should be compared to the rival ones and the statistics should show agreement rather than association,
11) Each method should be tested also by others than the original authors,
12) Authors should take responsibility for the long-term development of the method and take care of the usually necessary revisions.

We recommend that persons planning and implementing health surveys get familiar with:

1) the checklist by McDowell & Newell and

2) the APA standards.

For recommendations on mental health, see Lehtinen et al. 2001.

In a short run measures should be used, which have been widely applied in health surveys or preferably have characteristics of a standard measure of the target construct. To ensure comparability, these measures should as a rule not to be adapted but instead the original scales should be used. However, development work is essential. Therefore, it would be useful to validate new or modified scales, i.e. to include the scale under development work among older validated measures in a sample and investigate its characteristics. In the long run, international comparability can only be improved by development of new mental health measures for health survey purposes. Such work should be carried out in international collaboration, and include thorough testing of validity and reliability in a number of countries.

REFERENCES

Aalto, A.-M., Aro, A., & Teperi, J. (1999). RAND-36 terveyteen liittyvän elämänlaadun mittarina. Mittarin luotettavuus ja suomalaiset väestöarvot. (Stakes tutkimuksia 101). Saarijärvi: Stakes.

Andrews, G., & Peters, L. (1998). The psychometric properties of the Composite International Diagnostic Interview. Social Psychiatry & Psychiatric Epidemiology, 33, 80-88.

APA. (1985). Standards for psychological testing. Washington DC: American Psychological Association.

APA. (1999). Standards for educational and psychological testing. Washington DC: American Educational Research Association.

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale--I. Item selection and cross- validation of the factor structure. J Psychosom Res, 38(1), 23-32.

Bagby, R. M., Taylor, G. J., & Parker, J. D. (1994). The Twenty-item Toronto Alexithymia Scale--II. Convergent, discriminant, and concurrent validity. J Psychosom Res, 38(1), 33-40.

Bagby, R. M., Taylor, G. J., Parker, J. D., & Loiselle, C. (1990). Cross-validation of the factor structure of the Toronto Alexithymia Scale. J Psychosom Res, 34(1), 47-51.

Beck, A., Steer, R., & Garbin, M. (1988). Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. Clinical Psychology Review, 8(77-100).

Bhui, K., Bhugra, D., & Goldberg, D. (2000). Cross-cultural validity of the Amritsar Depression Inventory and the General Health Questionnaire amongst English and Punjabi primary care attenders. Soc Psychiatry Psychiatr Epidemiol, 35(6), 248-254.

Bullinger, M., Alonso, J., Apolone, G., Leplege, A., Sullivan, M., Wood-Dauphinee, S., Gandek, B., Wagner, A., Aaronson, N., Bech, P., Fukuhara, S., Kaasa, S., & Ware, J. E., Jr. (1998). Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol, 51(11), 913-923.

Cohen, S. R., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. Psychological Bulletin, 98, 310-357.

Cottler, L. B., Grant, B. F., Blaine, J., Mavreas, V., Pull, C., Hasin, D., Compton, W. M., Rubio-Stipec, M., & Mager, D. (1997). Concordance of DSM-IV alcohol and drug use disorder criteria and diagnoses as measured by AUDADIS-ADR, CIDI and SCAN. Drug Alcohol Depend, 47(3), 195-205.

Cronbach, L. J. (1990). Essentials of psychological testing (5. ed.). New York: HarperCollinsPublishers Inc.

Derogatis, L., & Cleary, P. (1977). Confirmation of the dimensional structure of the SCL-90: A study in construct validation. Journal of Clinical Psychology, 33(4), 981-989.

Derogatis, L., & DellaPietra, L. (1994). Psychological tests in screening for psychiatric disorder. In M. Maruish (Ed.), The use of psychological testing for treatment planning and outcome assessment (22-54). Hillsdale, USA: LEA.

Derogatis, L., & Melisaratos, N. (1983). The Brief Symptom Inventory: an introductory report. Psychological medicine, 13, 595-605.

Derogatis, L. R., & Lazarus, L. (1994). SCL-90-R, Brief Symptom Inventory, and matching clinical rating scales. In M. Maruish (Ed.), The use of psychological testing for treatment planning and outcome assessment (217-248). Hillsdale, USA: LEA.

Dunkel-Schetter, C., Folkman, S., & Lazarus, R. S. (1987). Correlates of social support receipt. Journal of Personality and Social Psychology, 53, 71-80.

ECHI Project. (2001). Design for a Set of European Community Health Indicators (Final report): group under National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands.

Fink, P., Ewald, H., Jensen, J., Sörensen, L., Engberg, M., Holm, M., & Munk-Jörgensen, P. (1999). Screening for somatization and hypochondriasis in primary care and neurological in-patients: a seven item scale for hypochondriasis and somatization. Journal of Psychosomatic Reserach, 46(3), 261-273.

Gandek, B., Ware, J. E., Jr., Aaronson, N. K., Apolone, G., Bjorner, J. B., Brazier, J. E., Bullinger, M., Kaasa, S., Leplege, A., Prieto, L., & Sullivan, M. (1998). Cross-Validation of Item Selection and Scoring for the SF-12 Health Survey in Nine Countries: Results from the IQOLA Project. J Clin Epidemiol, 51(11), 1171-1178.

Gandek, B., Ware, J. E., Jr., & the IQOLA Project Group. (1998). Methods for validating and norming translations of health status questionnaires: The IQOLA Project approach. J Clin Epidemiol, 51, 953-959.

Gardos, G., Schniebolk, S., Mirin, S., Wolk, P., & Rosenthal, K.-L. (1984). Alexithymia: Towards validation and measurement. Comprehensive Psychiatry, 25(3), 278-282.

Gatz, M., Johansson, B., Pedersen, N., Berg, S., & C, R. (1993). A cross-national self-report measure of depressive symptomatology. International Psychogeriatrics, 5(2), 147-156.

Goldberg, D., Gater, R., Sartorius, N., & al., e. (1997). The validity of the two versions of the GHQ in the WHO study of mental illness in general health care. Pychol Med, 27, 191-197.

Hibbett, M., Jagger, C., Polge, C., Cambois, E., & Ritchie, K. (1999). Cross European mental health indicators. A dream or reality? Eur J Pub Health, 9, 285-289.

HIS/HES database. (2000). HIS and HIS/HES Surveys in EU. Statistic Netherlands.

HIS/HES database. (2001, October 2001). HIS and HIS/HES Surveys in EU (2.). Scientific Institute of Public Health, Belgium and National Public Health Institute KTL, Finland.

HIS/HES database. (2002, January 2002). European Health Surveys DATABASE (3.). C.Vermeire & J.Tafforeau, Scientific Institute of Public Health, Belgium and National Public Health Institute KTL, Finland.

Holi, M., Sammallahti, P., & Aalberg, V. (1998). A Finnish validation study of the SCL-90. Acta Psychiatrica Scandinavica, 97, 42-46.

House, J. S., Landis, K. R., & Umberson, D. (1988). Social relationships and health. Science, 241, 540-545.

Hupkens, C., & Swinkels, H. (2001). Health interview surveys in the European Union: Overview of methods and contents. Health Surveys in the EU: HIS and HIS/HES evaluations and models, report from phase 1.: Statistics Netherlands.

Jagger, C., & Robine, J. (2001). Euro-REVES: Setting up a coherent set of health indicators: Report on Consensus workshop on Mental health, Montpellier 16/3/2001. Montpellier: Euro-REVES.

Joukamaa, M., Miettunen, J., Kokkonen, P., Koskinen, M., Julkunen, J., Kauhanen, J., Jokelainen, J., Veijola, J., Läksy, K., & Järvelin, M.-R. (2001). Psychometric properties of the Finnish 20-item Toronto Alexithymia Scale. Nord J Psychiatry, 55, 123-127.

Journal of Clinical Epidemiology, Special Issue on SF-36 measure. (1998). 51(11).

Katz, R. C., Katz, J. N., & Shaw, B. (1994). Beck Depression Inventory and Hopelessness Scale. In M. Maruish (Ed.), The use of psychological testing for treatment planning and outcome assessment (279-291). Hillsdale, USA: LEA.

Kendler, K., Gallagher, T., Abelson, J., & Kessler, R. (1996). Lifetime prevalence, demographic risk factors, and diagnostic validity of nonaffective psychosis as assessed in a US community sample. Archives of General Psychiatry, 53, 1022-1031.

Kessler, R. C. (1999). The World Health Organization International Consortium in Psychiatric epidemiology (ICPE): initial work and future directions - the NAPE lecture. Acta Psychiatr Scand, 99, 2-9.

Kessler, R. C., Andrews, G., Mroczek, D., Ustun, B., & Wittchen, H.-U. (1998). The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI-SF). International Journal of Methods in Psychiatric Research, 7(4), 171-185.

Koponen, P., & Aromaa, A. (2001). Health Examination Surveys (HES). Review of literature and inventory of surveys in the EU/EFTA Member States. Health Surveys in the EU: HIS and HIS/HES evaluations and models, report from phase 1.: National Public Health Institute KTL.

Korten, A., & Henderson, S. (2000). The Australian National Survey of Mental Health and Well-Being: Common psychological symptoms and disablement. British Journal of Psychiatry, 177, 325-330.

Lehtinen, V. (2002, 13-14 May 2002). Indicators for Mental Health Monitoring in Europe. Paper presented at the Health Surveys in Europe - Role of Surveys in Monitoring, Luxembourg.

Lehtinen, V., Korkeila, J., Morgan, A., Salize, H.-J., Dalgard, O. S., Bijl, R., Faria, J. S., & 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.

McDowell, I., & Newell, C. (1996). Measuring health. A guide to rating scales and questionnaires (2. ed.). New York: Oxford University Press.

Mui, A., Burnette, D., & Chen, L. (2001). Cross-Cultural Assessment of Geriatric Depression: A review of the CES-D and the GDS. Journal of Mental Health and Aging, 7, 137-163.

O'Reilly, P. (1988). Methodological issues in social support and social network research. Social Science and Medicine, 26, 863-873.

Parker, J., Bagby, R., Taylor, G., Endler, N., & P., S. (1993). Factorial validity of the 20-item Toronto Alexithymia Scale. European Journal of Personality, 7, 221-232.

Peters, L., & Andrews, G. (1995). Procedural validity of the computerized version of the Composite International Diagnostic Interview (CIDI-Auto) in the anxiety disorders. Psychological Medicine, 25, 1269-1280.

Pevalin, D. (2000). Multiple applications of the GHQ-12 in a general population samle: An investigation of long-term retest effects. Social Psychiatry & Psychiatric Epidemiology, 35, 508-512.

Pilowsky, I. (1967). Dimensions of hypochondriasis. Brit. J. Psychiat., 113, 89-93.

Prosser, D., Johnson, S., Kuipers, E., Dunn, G., Szmukler, G., Reid, Y., Bebbington, P., & Thornicroft, G. (1999). Mental health, "burnout" and jobsatisfaction in a longitudinal study of mental health staff. Social Psychiatry & Psychiatric Epidemiology, 34, 295-300.

Radloff, L. (1977). The CES-D Scale: A self-report depression scale for research in the general polulation. Applied Psychological Measurement, 1(3), 385-401.

Robins, L. (1989). Diagnostic grammar and assessment: translating criteria into questions. Psychological Medicine, 19, 57-68.

Schmitz, N., Kruse, J., Heckrath, C., Alberti, L., & Tress, W. (1999). Diagnosing mental disorders in primary care: the General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments. Social Psychiatry & Psychiatric Epidemiology, 34, 360-366.

Schutte, N., Toppinen, S., Kalimo, R., & Schaufeli, W. (2001). The factorial validity of the Maslach Burnout Inventory - General Survey (MBI-GS) across occupational groups and nations. Journal of Occupational and Organizational Psychology, 73, 53-66.

Spitzer, R., Williams, J., Gibbon, M., & First, M. (1992). The Structured Clinical Interview for DSM-III-R (SCID), I: History, rationale, and description. Archives in General Psychiatry, 49, 624-629.

Sullivan, M., Karsson, J., & Ware, J. J. (1995). The Swedish SF-36 Health Survey - I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general polulation in Sweden. Social Science and Medicine, 41(10), 1349-1358.

Taylor, G., Bagby, R., Ryan, D., Parker, J., Doody, K., & Keefe, P. (1988). Criterion validity of the Toronto Alexithymia Scale. Psychosomatic Medicine, 50, 500-509.

The HISHES Group. (2003). European Health Interview & Health Examination Surveys Database. CBS, Netherlands, IPH, Belgium & KTL, Finland. Available: https://**www.iph.fgov.be/hishes/** . [2003, 26. June 2003].

The WHOQOL Group. (1995). The World Health Organization Quality of Life Assessment (WHOQOL): position paper from the World Health Organization. Social Science and Medicine, 41, 1403-1409.

The WHOQOL Group. (1998a). Development of the World Health Organization WHOQOL-BREF Quality of Life Assessment. Psychological Medicine, 28, 551-558.

The WHOQOL Group. (1998b). The World Health Organization Quality of Life assessment (WHOQOL): Development and general psychometric properties. Social Science and Medicine, 46(12), 1569-1585.

Ware, J. J., Gandek, B., & for the IOQLA Project Group. (1998). Overview of the SF-36 Health Survey and the International Quality of Life Assessment IQOLA) Project. J Clin Epidemiol, 51, 903-912.

WHO. (1997). Composite International Diagnostic Interview (CIDI, version 2.1). Geneva: World Health Organization.

WHO. (1999). The World Health Report 1999, Making a Difference. Geneva: World Health Organization.

WHO. (2001). The World Health Report 2001, Mental health: New understanding, new hope. Geneva: World Health Organization.

WHOQOL Group. (1994). Development of the WHOQOL: rationale and current status. International Journal of Mental Health, 23, 24-56.

Wittchen, H.-U. (1994). Reliability and validity studies of the WHO-composite international diagnostic interview (CIDI): A critical review. J Psychiat Res, 28, 57-84.

Wittchen, H.-U., Lachner, G., Wunderlich, U., & Pfister, H. (1998). Test-retest reliability of the computerized DSM-IV version of the Munich-Composite International Diagnostic Interview (M-CIDI). Social Psychiatry & Psychiatric Epidemiology, 33, 568-578.

Appendix 1 Mailed questionnaire on the use of mental health and quality of life measures in national HIS and HES surveys.

Dear Sir/Madam,

As a subproject of the EU project 'Health Surveys in the EU: HIS and HIS/HES evaluations and models', phase 2, we investigate measurement on mental health issues obtained by validated questionnaires:

Mental health measurement in national health interview (HIS) and health examination surveys (HES).

Your Survey

_____

may contain some measures relevant to this respect. In order to find out more, we kindly ask you to provide us additional information about the

1) **mental health** related and
2) **quality of life**

instruments, questionnaires and scales (later indicated as **measures**) used in your survey.

1. In the HIS/HES database, the following mental health and quality of life measures were indicated in your survey.

Name of measure	Abbreviated name	In original form	In adapted form
1.	_____	☐	☐
2.	_____	☐	☐
3.	_____	☐	☐
4.	_____	☐	☐
5.	_____	☐	☐
6.	_____	☐	☐
7.	_____	☐	☐
8.	_____	☐	☐

2. Check the list above
    a) the exact name and version (e.g., number of items) of the measure(s) used, and

    b) the correct abbreviation(s) of each measure (especially the use of capital letters).

3. a) Did you adapt or modify the measure(s) for the purposes of your survey?

    No   ☐   Yes   ☐

If you did

1. please check the list above and indicate which measures you modified,

2. describe below the modifications you performed

    _____

    _____

    _____

    _____

    b) Did you use any measure(s) only in part?
    No   ☐   Yes   ☐

3. indicate which parts (e.g., which subscales) of a measure you used.

    _____

    _____

    _____

4. Did you use any of the measure(s) only in a subsample of your survey?
    No ☐   Yes   ☐

    If *Yes*, indicate below these measures and the parts of the survey the measures were used in (give, e.g., age limits or other inclusion or exclusion criteria).

    _____

    _____

    _____

5. Was any of the measure(s) a copyrighted one?
    No ☐   Yes   ☐   I don't know ☐

    If *Yes*, indicate below these measures, copyright holder of the measure, and information on the permission to use it in your survey.

    _____

    _____

6. Was any of the replies entered directly into computer (a computer generated measure)?

No ☐   Yes   ☐

If *Yes*, indicate below these measure(s) and information on the computer program used.

_____

_____

7. Did your survey/project develop mental health related or quality of life measure(s) for the purposes of this survey?

No ☐   Yes   ☐

If *Yes*, indicate below

   a) these measures

   _____

   _____

   b) the person(s) in charge of the developing work, and contact information or references (publications) if available.

   _____

   _____

   _____

8. The next section (Appendices A – X) contains detailed information on the following measures

   –Describe each measure in a separate appendix.

   Name of measure   Appendix attached

   - _____   ☐   Appendix A
   - _____   ☐   Appendix B
   - _____   ☐   Appendix C
   - _____   ☐   Appendix D
   - _____   ☐   Appendix E
   - _____   ☐   Appendix F
   - _____   ☐   Appendix G

Appendix   A – X

Name of the measure _____

A 1. Original language of the measure

	Yes	No
a)The measure was an international one, that is, available simultaneously in several languages, and one of these languages was suitable for use in your survey.	☐	☐
b) The measure was originally in English.	☐	☐

c) If the measure was not international or in English,
indicate here the original language(s) of the measure.

_____

_____

A 2. a) Did your survey or project translate this measure into one or several local
    language(s)?
    No ☐   Yes   ☐

    These languages: _____

   b) Have the measures been previously translated into local language(s)?

    No ☐   Yes   ☐

    If *Yes* in either of the previous questions, give information on the project/
    institute etc. in charge of the translation work.
    _____

    _____

A 3. Has the same measure (according to your knowledge) been used before in your
    country in
    a)   population surveys   No   ☐   Yes ☐

    b)   other population settings   No   ☐   Yes ☐

    c)   clinical settings   No   ☐   Yes ☐

    If *Yes* in any of the previous, indicate information on these surveys or research
    (name of the project etc.)_____

    _____

    _____

A 4. Has validity or reliability of the measure been investigated/ tested in a population sample in your country?

No ☐ Yes ☐

If *Yes*, give references – and/or name(s) and contact information of person(s) – with knowledge of this validation work.

_____

_____

A 5. Give main reason(s) for selecting this measure for your national health survey.

                                                                            Yes

- Measure was applied earlier in local or national surveys(s) of the country ☐

- Measure was applied earlier in relevant national survey(s) in other ☐

  countries (for comparability)

- An internal expert group of your survey recommended this measure ☐

- An external expert group recommended this measure ☐

- The survey group has been working with this measure before ☐

- Measure were easy to apply

  a) for the subjects ☐

  b) for the survey (costs, etc.) ☐

- For other reasons, indicate below

_____

_____

_____

A 6. Indicate some literature reference(s), which were important for you when selecting or applying the measure.

_____

_____

_____

9. If possible, **attach the measure(s)** in the form they were used in your survey

    a) in the language they were used and

    b) also in English, if available.

If you have questions, please don't hesitate to contact Ulla-Sisko Lehto-Järnstedt.

Indicate below your (the respondent's) name and contact information.

_____

_____

_____

Thank you for your help and for your time!

Ulla-Sisko Lehto-Järnstedt   Arpo Aromaa

Senior Researcher   Professor

National Public Health Institute KTL

Department of Health and Disability

Mannerheimintie 166

FIN-00300 Helsinki

Phone 358 9 4744 8795

Fax   358 9 4744 8924

ulla-sisko.lehto-jarnstedt@ktl.fi