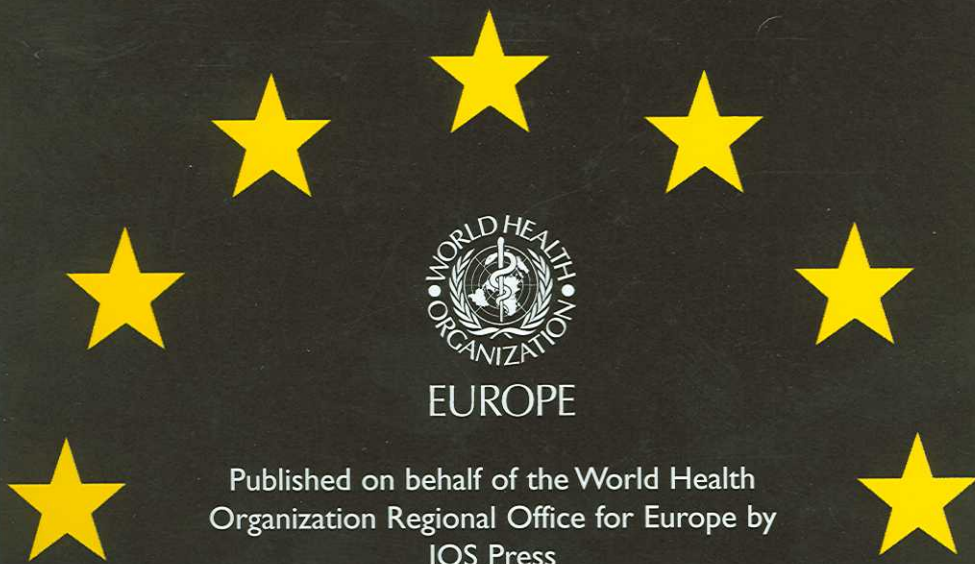




EUROHIS

DEVELOPING COMMON INSTRUMENTS FOR HEALTH SURVEYS

Editors:
Anatoliy Nosikov
Claire Gudex



Published on behalf of the World Health
Organization Regional Office for Europe by
IOS Press

*Biomedical and
Health Research*



IOS
Press

OHM
Ohmsha

Biomedical and Health Research

Volume 57

Recently published in this series:

- Vol. 56. P. Chauvin and the Europromed Working Group (Eds.), *Prevention and Health Promotion for the Excluded and the Destitute in Europe*
- Vol. 55. J. Matsoukas and T. Mavromoustakos (Eds.), *Drug Discovery and Design: Medical Aspects*
- Vol. 54. I.M. Shapiro, B.D. Boyan and H.C. Anderson (Eds.), *The Growth Plate*
- Vol. 53. C. Huttin (Ed.), *Patient Charges and Decision Making Behaviours of Consumers and Physicians*
- Vol. 52. J.-F. Stoltz (Ed.), *Mechanobiology: Cartilage and Chondrocyte, Vol. 2*
- Vol. 51. G. Lebeer (Ed.), *Ethical Function in Hospital Ethics Committees*
- Vol. 50. R. Busse, M. Wismar and P.C. Berman (Eds.), *The European Union and Health Services*
- Vol. 49. T. Reilly (Ed.), *Musculoskeletal Disorders in Health-Related Occupations*
- Vol. 48. H. ten Have and R. Janssens (Eds.), *Palliative Care in Europe – Concepts and Policies*
- Vol. 47. H. Aldskogius and J. Fraher (Eds.), *Glial Interfaces in the Nervous System – Role in Repair and Plasticity*
- Vol. 46. I. Philp (Ed.), *Family Care of Older People in Europe*
- Vol. 45. L. Turski, D.D. Schoepp and E.A. Cavalheiro (Eds.), *Excitatory Amino Acids: Ten Years Later*
- Vol. 44. R. Coppo and Dr. L. Peruzzi (Eds.), *Moderately Proteinuric IgA Nephropathy in the Young*
- Vol. 43. B. Shaw, G. Semb, P. Nelson, V. Brattström, K. Mølsted and B. Prahll-Andersen, *The Eurocleft Project 1996-2000*
- Vol. 42. J.-F. Stoltz (Ed.), *Mechanobiology: Cartilage and Chondrocyte*
- Vol. 41. T.M. Gress (Ed.), *Molecular Pathogenesis of Pancreatic Cancer*
- Vol. 40. H. Leino-Kilpi, M. Välimäki, M. Arndt, T. Dassen, M. Gasull, C. Lemonidou, P.A. Scott, G. Bansemir, E. Cabrera, H. Papaevangelou and J. Mc Parland, *Patient's Autonomy, Privacy and Informed Consent*
- Vol. 39. J.-M. Graf von der Schulenburg (Ed.), *The Influence of Economic Evaluation Studies on Health Care Decision-Making*
- Vol. 38. N. Yoganandan and F.A. Pintar (Eds.), *Frontiers in Whiplash Trauma*
- Vol. 37. J.M. Ntambi (Ed.), *Adipocyte Biology and Hormone Signaling*
- Vol. 36. F.F. Parl, *Estrogens, Estrogen Receptor and Breast Cancer*
- Vol. 35. M. Schlaud (Ed.), *Comparison and Harmonisation of Denominator Data for Primary Health Care Research in Countries of the European Community*
- Vol. 34. G.J. Bellingan and G.J. Laurent (Eds.), *Acute Lung Injury: From Inflammation to Repair*
- Vol. 33. H.H. Goebel, S.E. Mole and B.D. Lake (Eds.), *The Neuronal Ceroid Lipofuscinoses (Batten Disease)*
- Vol. 32. B.J. Njio, B. Prahll-Andersen, G. ter Heege, A. Stenvik and R.S. Ireland (Eds.), *Quality of Orthodontic Care – A Concept for Collaboration and Responsibilities*
- Vol. 31. B.J. Njio, A. Stenvik, R.S. Ireland and B. Prahll-Andersen (Eds.), *EURO-QUAL*
- Vol. 30. J.-F. Stoltz, M. Singh and P. Riha, *Hemorheology in Practice*
- Vol. 29. G. Pawelec (Ed.), *EUCAMBIS: Immunology and Ageing in Europe*
- Vol. 28. A.M.N. Gardner and R.H. Fox, *The Venous System in Health and Disease*
- Vol. 27. P.A. Frey and D.B. Northrop (Eds.), *Enzymatic Mechanisms*
- Vol. 26. P. Jenner (Ed.), *A Molecular Biology Approach to Parkinson's Disease*
- Vol. 25. N.J. Gooderham (Ed.), *Drug Metabolism: Towards the Next Millennium*
- Vol. 24. S.S. Baig (Ed.), *Cancer Research Supported under BIOMED I*
- Vol. 23. M. Hallen (Ed.), *Human Genome Analysis*
- Vol. 22. J. Matsoukas and T. Mavromoustakos (Eds.), *Bioactive Peptides in Drug Discovery and Design: Medical Aspects*

EUROHIS: Developing Common Instruments for Health Surveys

Edited by

Anatoliy Nosikov

*World Health Organization Regional Office for Europe, Copenhagen,
Denmark*

and

Claire Gudex

*World Health Organization Regional Office for Europe, Copenhagen,
Denmark*



Amsterdam • Berlin • Oxford • Tokyo • Washington, DC

Published on behalf of the World Health Organization Regional Office for Europe by IOS Press

© 2003, World Health Organization

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without prior written permission from the publisher.

ISBN 1 58603 322 0 (IOS Press)

ISBN 4 274 90579 9 C3047 (Ohmsha)

Library of Congress Control Number: 2003 100033

Publisher

IOS Press

Nieuwe Hemweg 6B

1013 BG Amsterdam

The Netherlands

fax: +31 20 620 3419

e-mail: order@iospress.nl

Distributor in the UK and Ireland

IOS Press/Lavis Marketing

73 Lime Walk

Headington

Oxford OX3 7AD

England

fax: +44 1865 75 0079

Distributor in the USA and Canada

IOS Press, Inc.

5795-G Burke Centre Parkway

Burke, VA 22015

USA

fax: +1 703 323 3668

e-mail: iosbooks@iospress.com

Distributor in Germany, Austria and Switzerland

IOS Press/LSL.de

Gerichtsweg 28

D-04103 Leipzig

Germany

fax: +49 341 995 4255

Distributor in Japan

Ohmsha, Ltd.

3-1 Kanda Nishiki-cho

Chiyoda-ku, Tokyo 101-8460

Japan

fax: +81 3 3233 2426

LEGAL NOTICE

The publisher is not responsible for the use which might be made of the following information.

PRINTED IN THE NETHERLANDS

Foreword

The availability and quality of data from population health surveys in Europe have greatly improved over the last 10–15 years, particularly in the countries of central and eastern Europe. The survey approach is now fully recognized to be a valuable method for health monitoring that is complementary to the registration approach. Nevertheless, the comparability of health survey data remains a challenge. Difficulties arise from two main sources: “man-made” variations between surveys owing to methodological differences (for example, in the way that indicators and health conditions are defined) and “natural” differences between populations caused by varying attitudes, behaviour and concepts of health. Moreover, it is not always possible to reliably identify and separate the two.

The establishment of agreed standards in terms of recommended common methods and measurement instruments can largely improve control of the first main source of incomparability and, at the same time, increase the relevance and reliability of survey data. This has immediate benefits for public health practice, as we continuously make international comparisons to provide evidence for health policies. In the long term, however, we must reach beyond international harmonization of instruments: we must greatly improve the scientific understanding of the causes of any significant incomparability of health surveys between countries.

These tasks are impossible without sincere international collaboration, with careful consideration of the similarities and differences among health concepts, approaches and patterns in different countries and organizations. The knowledge derived from such projects should be used to customize health strategies, models and skills so that they are better tailored to the specific needs and resources of each country. The special added value of such research arises from the cross-fertilization of ideas between cultures (including organizational cultures), the facilitation of multidisciplinary research, and the creation of links between health research, health policy and the individual citizen. The EUROHIS project has capitalized on this thinking.

The project has required financial and scientific input from many different sources as well as a carefully planned programme of work. In particular, the support and generous financial contribution of the European Commission’s Biomedical and Health Research Programme (BIOMED 2), which has financed the project as a Concerted Action, is gratefully acknowledged. EUROHIS demonstrates how WHO and the Commission can work collaboratively to deliver results that have greater impact than if the organizations worked alone. The project has contributed to the objectives of both organizations. It has allowed greater cross-cultural diversity and has therefore produced results of more general validity and relevance. The success of the EUROHIS project is clear, but the full impact of it will be decided by the use of its recommendations by the public health authorities in the Member States.



Anca Dumitrescu
Director, Division of Information,
Evidence and Communication,
WHO Regional Office for Europe



Octavi Quintana Trias
Director, Health Research,
Commission of European Communities

Acknowledgements

EUROHIS is a joint project of the Regional Office and the European Commission's BIOMED2 programme. The generous financial support of the Commission is gratefully acknowledged.

The project was directed by a Scientific Board comprising Arun Nanda and subsequently Anatoliy Nosikov (Chair), together with Viviana Egidi, Martti Heikkinen, Howard Meltzer, Jussi Simpura, Pekka Oja, Jaap van den Berg, Agnes de Bruin, Esko Kalimo, Timo Klaukka, Marijke de Kleijn-de Vrankrijker and Mick Power (Principal Investigators). The overall coordination and management of the project was provided by Anatoliy Nosikov and Claire Gudex, while the Principal Investigators were responsible for the management of the indicator-specific networks. Valuable input was received from Arpo Aromaa and Jean-Marie Robine (advisers to the Board), as well as from Jakob B. Bjørner, who assisted the quality of life network at the beginning of the project. Milagros Garcia Barbero, WHO European Office for Integrated Health Care Services, Barcelona, assisted the use of curative medical services network.

The general concept of the project was formulated by Arun Nanda, Anatoliy Nosikov and Esko Kalimo. The specific concepts and methods for the individual indicators were developed by the Principal Investigators and the active participants in the indicator networks. The EUROHIS indicator networks are listed in Annex 1. The indicator-specific approaches and the individual steps were discussed at international consultative meetings. All the participants in the project are listed in Annex 2. The active participants in the indicator networks also contributed to the drafting and review of the manuscript of the book.

The protocols for the translation, the pre-testing and the field-testing were developed by Dirkjan Beukenhorst, Monika Bullinger and Claire Gudex, with input from the Principal Investigators. Dirkjan Beukenhorst also provided analysis of the pre-test data. The field-test translations into French, German and Russian were carried out by the translation services at the Regional Office. The field-test translations into the other languages were organized by the national counterparts. Viviane Kovess, Monika Bullinger, Thomas Ziese, Tamara Maximova, Olga Gaenko and Vitaliy Belov provided comments on forward translations, and the back translations were reviewed by Claire Gudex.

The field-test data were collected by the market research institute, NIPO (Amsterdam, the Netherlands), in France, Germany and the United Kingdom, and by working groups in Azerbaijan, Croatia, the Czech Republic, Israel, Kyrgyzstan, Latvia, Lithuania, the Republic of Moldova, Romania, the Russian Federation, Slovakia and the former Yugoslav Republic of Macedonia. Technical input to the data processing was provided by Abdonas Tamosiunas and Kristina Jureniene, Monika Bullinger, Silke Schmidt and Lars Moeller.

Administrative and technical support for the project was provided by Marie Kruse, Birthe Rasmussen and other WHO staff. Language editing was by Frank Theakston, and the book was designed and laid out by IOS Press.

Introduction

The adoption by WHO in 1977 of the health for all strategy highlighted the need for internationally comparable health data derived from population surveys. Such data can identify true variations in health status and the factors that affect health. Policy discussions on best practices to improve the health of populations can then be based on appropriate cross-national comparisons. However, the current lack of coordination in the development of survey instruments hinders international comparison of the results. The primary objective of the EUROHIS project was therefore to develop common survey instruments that could enhance the international comparability of national health data. The project, led by the WHO Regional Office for Europe, has spanned six years and involved the collaboration of researchers and survey managers from 33 countries in the WHO European Region.

This book describes the approach taken by the EUROHIS project and the methods used to address the various challenges that arose during the course of the project. It provides an account of the EUROHIS programme of work and the resulting recommendations for common instruments, and includes methodological discussions of cross-national comparability issues.

The EUROHIS project has culminated in a set of recommended common instruments for eight health for all indicators:

- chronic physical conditions
- mental health
- alcohol consumption
- physical activity
- use of curative medical services
- use of medicines
- use of preventive health care
- quality of life.

The instruments were developed in English and translations are available in French, German and Russian as well as 12 other languages: Arabic, Azerbaijani, Croatian, Czech, Hebrew, Kyrgyz, Latvian, Lithuanian, Macedonian, Romanian, Slovak and Ukrainian. Currently, several European Member States have either already applied the recommended instruments or are in the process of implementing them in their national surveys.

Structure of the book

Chapter 1 provides an overview of the methodological challenges that face international comparability studies and reports on current European efforts to harmonize survey instruments and methodology. Chapter 2 gives a more detailed description of the EUROHIS project, describing the general approach adopted and the main stages of work. Chapters 3–10 report on the work carried out on each of the eight indicators, including the methodologies used to develop the common instruments, the performance of these instruments in the field and recommendations for their use. Chapter 11 concludes the book.

Contents

Foreword	
<i>Anca Dumitrescu and Octavi Quintana Trias</i>	v
Acknowledgements	vi
Introduction	vii
1. International comparability of health interview surveys: An overview of methods and approaches <i>Monika Bullinger</i>	1
2. Overview of the EUROHIS project <i>Anatoliy Nosikov and Claire Gudex</i>	13
3. Development of a common instrument for chronic physical conditions <i>Vittoria Buratta, Luisa Frova, Lidia Gargiulo, Emilio Gianicolo, Sabrina Prati and Luciana Quattrociochi</i>	21
4. Development of a common instrument for mental health <i>Howard Meltzer</i>	35
5. Development of a common instrument for alcohol consumption <i>Jussi Simpura</i>	61
6. Development of a common instrument for physical activity <i>Pekka Oja</i>	79
7. Development of a common instrument for use of curative medical services <i>Agnes de Bruin and Dirkjan Beukenhorst</i>	93
8. Development of a common instrument for use of medicines <i>Timo Klaukka and Jaana Martikainen</i>	121
9. Development of a common instrument for use of preventive health care <i>Marijke W. de Kleijn-de Vrankrijker and Wim Davidse</i>	129
10. Development of a common instrument for quality of life <i>Mick Power</i>	145
11. Conclusions	165
Annex 1 EUROHIS indicator networks	169
Annex 2 List of EUROHIS project participants	179

Chapter 1.

International comparability of health interview surveys: An overview of methods and approaches

Monika Bullinger¹

Department of Medical Psychology, University Hospital of Hamburg Eppendorf, Germany

1.1 Introduction

The availability of representative population-based health data is a major prerequisite for identifying and understanding a community's public health problems, as well as for planning and evaluating health care options. These data are also needed for cross-national studies that aim to compare the provision of health services and the level of health in different populations (Murray & Lopez 1995; Ovretveit 2001; Zwi & Yach 2002). At the European level, the political impetus for such cross-national comparisons comes from the European unification process, which calls for common standards of documentation and intervention in several areas of public concern, including health-related issues (Donnan 1998; Borrell et al. 2000). A well known example at the global level is the *The world health report 2000* (WHO 2000).

There have always been diverse opinions on the measurement of health in epidemiology and the clinical sciences. While it is accepted that health is a theoretical construct that needs to be operationalized before it can be measured, approaches differ with regard to the aspects of health that are measured and place different emphasis on the "objective" and "subjective" aspects. The history of health indicator research shows that the priorities in this field have changed from macro-social structural information (e.g. infant mortality), through clinical indicators (e.g. blood cholesterol level), to self-reported health information (e.g. perceived health and wellbeing).

This change in emphasis towards the inclusion of self-reported aspects of health has its roots in the 1948 WHO definition of health (WHO 1985), which caused a paradigm shift in the conceptualization of health. The WHO definition encompasses not only physical but also mental and social wellbeing, and maintains that health should not simply be equated with absence of disease or infirmity. Although the WHO definition has been criticized for its idealized view of health, it has been influential in stressing the psychological and social dimensions of the health concept, as well as in focusing on the value of self-reported health. Nevertheless, WHO still focuses on more objective indicators in its statistical information system (WHO SIS) and in other major projects, such as the Global Burden of Disease studies.

¹ The views expressed in this chapter are those of the author and do not necessarily reflect the policies or opinion of the WHO Regional Office for Europe.

Much of the current interest centres around the differences that emerge between clinical measurements of health status and self-reported measurements of health, as well as the specific problems associated with each approach. There is uncertainty, for example, over the value of clinical laboratory data in the prognosis of future morbidity, while self-reported health data are influenced by psychological and methodological processes that make it difficult to assess their reliability and validity. Conceptually, subjective or “perceived” health is closely related to a relatively new area of research, namely “quality of life”. Quality of life can best be understood as the operationalization of the WHO definition of health, through the inclusion of self-report data on wellbeing and the level of functioning in physical, emotional and social domains of life (Bullinger 1991). Usual notions of quality of life emphasize the role of perceptions, i.e. cognitive and emotional evaluations of health-related information, rather than simply self-reported behaviour and function.

In the current context, it is useful to distinguish between “health status” (i.e. the self-reported *description* of health as related to symptoms, behaviour and experiences) and “quality of life” (i.e. the self-reported *evaluation* of symptoms, behaviour and experiences). The conceptual overlap between self-reported health status and quality of life is difficult to eliminate, however. This is apparent, for example, in the original operationalization of health-related quality of life by the WHOQOL group (Orley & Kuyken 1994), who defined quality of life as the “extent to which individual expectations are met by experience”, and measured it not only by self-reported evaluations (“am satisfied with”) but also by self-reported functions (e.g. “can do”).

The search for relationships between objective and subjective health indicators reveals that, depending on the nature of the symptoms and the diagnosis, individuals can be well informed about their health status and can reliably report symptoms and diagnosis in accordance with clinical information (Stewart & Ware 1992). However, as is apparent in the quality of life literature, the evaluation of symptoms, experiences and behaviour is neither directly nor proportionally related to clinical indicators such as the severity of symptoms. To fully capture the concept of health status, therefore, its measurement should be operationalized via clinical information and self-reported data (which is in line with the classical approach to health indicator research; Cavelaars et al. 1998) and complemented by indicators of self-perceived quality of life.

The issue of cross-cultural or international comparability of health data is critical in view of the increasing globalization of health. Not only is health affected by economic globalization, it is itself an object for globalization owing to its cross-cultural nature (Cornia 2001). The main challenges that face assessment of health indicators are (a) whether health indicators do indeed tap relevant constructs, (b) whether they can be operationalized into survey instruments in such a way that enables individuals to respond to them, and (c) whether the survey instruments are methodologically robust. Cross-cultural assessment faces the additional challenge (d) of whether the information that is obtained can be interpreted within and across cultures. In addition, the international harmonization of health indicators that are already in use in different countries requires the development of criteria by which their applicability in a cross-cultural context can be judged, as well as innovative statistical approaches for cross-cultural analysis.

Both top-down and bottom-up approaches have been suggested to derive cross-cultural health indicators. Top-down approaches are derived from a theoretical basis (e.g. via knowledge of health systems) while bottom-up approaches are derived from an empirical basis (e.g. by identifying frequently collected data) (Elola et al. 1995). A bottom-up approach would require domains of health to be identified by experts (or the general public) that are then used to select options for operationalization. The top-down approach would be to review the currently available health indicators and critically assess their applicability to different cultural contexts. A combination of the two strategies has been

suggested as the method of choice.

1.2 Relevant health indicators

A review of the literature on health indicators and international health reporting reveals a substantial variation in the number and type of indicators used in population surveys.

A recent literature search in the Medline system (1993–2002) using relevant search terms resulted in a surprising paucity of pertinent articles. Although 659 articles on “health indicators” combined with “international” were available, only 127 were related to “health indicators” and “Europe”. Thirty-nine papers were available on the search words “health indicators” and “cross-cultural”, while only eight were found on “representative surveys” and “Europe”. With the search term “health reporting and international”, only six papers were identified. One of these describes international efforts to measure the quality of life of children and adolescents within the European Union (Ravens-Sieberer et al. 2001). Another article, which reviews the collection of comprehensive health data and health information in Europe (Aromaa 1998), describes health survey projects involving WHO, the Organisation for Economic Co-operation and Development (OECD), the Centers for Disease Control and Prevention (CDC) and other international organizations that collect health data and publish health survey reports. Between 1995 and 1998, the European Union (EU) initiated several activities to improve health statistics and to develop health monitoring. Related to the EU public health programmes, the health monitoring programme (HMP) was started in 1997 with the intention of paving the way towards permanent health monitoring by the EU.

It is only recently that there has been an increased effort to harmonize instruments and study designs in epidemiological studies. In the mental health field, for example, a recent review of the data in Europe (Hupkens et al. 1999) identified 26 surveys conducted in 13 European countries that contained information on mental health. However, these surveys rarely used standard measurement instruments that would provide data within the International Classification of Diseases (ICD) or the Diagnostic and Statistical Manual (DSM) of the American Psychiatric Association. A recent review of health surveys in OECD countries (Gudex & Lafortune 2000) showed that over 50% of the national health surveys included at least some measures of psychological and emotional wellbeing but, again, differences in the instruments used seriously limited the comparability of the data collected.

Several European research groups are currently active in addressing cross-cultural aspects of health indicators. In the field of mental health, these include the mental health group of the health monitoring programme of the European Union (HMP), the mental health project led by the European Network on Health Expectancy, and the Disability Process (EURO-REVES) also funded through the HMP. As regards environmental health risks, the European Health and Environment Information System for Disease and Exposure Risk Assessment (EUROHEIS, www.met.ic.ac.uk) is active, as is the Regional Office for Europe of the World Health Organization, which has developed an environment and health information system (www.who.dk.eprise/main/who/Prog/EHI/home). More globally, CDC (www.cdc.gov/mmwr/indss_99.html) summarizes survey programmes for selected public health indicators within specific population groups (older, younger persons), specific conditions (AIDS, asthma, malaria) and diagnostic/therapeutic techniques (e.g. mammography). In addition, there is the European Working Group on Socio-Economic Inequalities in Health (www.kttl.helsinki.fi) as well as programmes that relate social indicators to health indicators, such as the poverty net, which monitors living conditions and their effect on health at the global level (www.worldbank.org/poverty/data).

The health indicators presented in public health reports typically include sociodemographic variables, incidence and prevalence rates of diseases and disabilities, health care utilization data and information on health behaviour. The inclusion of subjective wellbeing, social support, accidents and sexual issues is less common, while there is rarely valid data on health-related quality of life, life events, stress or violence. The European Health Survey Data Base, which is an EU-funded project, maintains a data bank of the European health surveys and includes information on content, aims, instruments, sample size and design (to be available on the Internet from January 2003). However, instruments that would enable international assessment of chronic conditions, physical, mental and social health or personal resources are largely lacking so far. In an attempt to harmonize the data that are collected in the national health surveys of different countries, an EU-initiated task force is currently assessing 12 health topics, with the aim of producing data that are statistically comparable across countries. Among these topics are chronic conditions, health behaviour, health care utilization and medication, as well as perceived subjective health or health-related quality of life. In parallel, the European Community Household Panel (ECHP) manages an annual multidimensional survey that includes items on income, demographics, health (including mental health), addiction, living conditions and migration. However, harmonization attempts have not been conducted systematically. A useful source of information on health reporting and monitoring in childhood and adolescence is a recent international study (Currie et al. 2001). This study, as well as other international approaches that have been applied in the child and adult area, has been a useful source of information in the preparation of the German child health survey (<http://www.kinder-jugend-gesundheit21.de/>).

It is clear that there is a wide variety of research being undertaken at both international and European levels with regard to health indicators. These studies range from one-off research projects that collect data about specific health problems, to ongoing collaborative efforts to harmonize the monitoring and reporting of health indicators, both in the environment/health area as well as within national health survey programmes (e.g. WHO-SIS, CDC or OECD data).

Nevertheless, the standardization of health indicators commonly used in national health reporting has only recently been addressed. Considering the wide variation in research questions, methods and approaches, such an effort is highly relevant. Within Europe, it would seem that national statistical offices (in addition to global approaches within WHO or OECD) offer the best opportunities to monitor population health over time. If it were possible to harmonize and standardize data collection within these agencies, a significant database would be available for planning and monitoring health policy in the EU. The specific objective of the EUROHIS project was to contribute to the establishment of such a database.

Interestingly, the current literature provides no evidence for the performance of health indicators across countries, with the exception of the mental health area, in which the ability of instruments to fulfil psychometric properties has been more frequently tested. Such information is necessary in the identification of methodologically robust instruments and to examine the specific contribution of the standardized health indicators in Europe. The EUROHIS project can therefore also serve as a pioneer with regard to testing the performance of health measures with regard to cross-population analysis.

1.3 Cross-cultural harmonization and use of health indicators

The issue of health monitoring and reporting has received increasing attention from within the public health field. A recent development within this field is an increasing demand for international health surveys, which provoke discussions about conceptual clarity, methodology and practical applications in the international health field.

Health survey data can be used both to describe the health status of populations (epidemiological perspective) and to assess the outcome of interventions within health systems (clinical perspective), thereby contributing to decision-making in the public health field (policy perspective). These objectives are relevant at both national and international levels.

The term “international” has different meanings: politically it refers to nations, geographically it refers to countries, anthropologically it refers to cultures, sociologically it refers to societies, and psychologically it refers to the identity of its members. All of these meanings are reflected in the notion of language, which makes this one of the key issues in working with health assessments internationally. It is helpful to distinguish between the terms “international” and “cross-cultural”. “International” primarily denotes joint activities of different countries in the health field (i.e. studies from different countries concerning specific conditions). “Cross-cultural” research denotes an additional collaborative and comparative effort to explain specific differences in the health field (i.e. using specific instrumentation to assess health conditions across cultures). More recently, the term “cross-cultural” has been replaced by the more neutral term “cross-population”.

From a cross-population perspective, an individual’s assessment of health may be largely determined by cultural schemata. As Hutchinson (1996) pointed out, the culture-specific meanings of disease are most clearly noticeable in the so-called “folk illnesses”. If disease, as anthropological research suggests, is so very much culture-bound, how could health be culture-free? The basic scepticism, especially of anthropology, is captured in the following quotation (Johnson 1996, p. 511):

Although some researchers may desire a scale or similar instruments for global assessments of cultures, permitting comparison of the ‘nature’ of one culture with that of another, no such scale exists. In fact, given the multiplicity of variables or domains comprising a culture, that goal is unrealistic, both theoretically and methodologically.

Given the increasing interest in cross-culturally applicable health measures, the following questions merit consideration.

- Which indicators of health are relevant in a given population or culture?
- Do population/cultural groups share identical sets of concepts about health?
- How can health indicators be assessed in a methodologically robust way (reliability, validity, responsiveness)?
- Can health be measured across populations/cultures using the same data collection procedures?
- Can data on health indicators be compared across populations/cultures?
- Do cross-cultural health data provide a sound basis for decision-making in the national and international health care fields?

Almost no literature can be found on the development of cross-culturally applicable health indicators, although papers are available on cross-cultural quality of life research (Anderson et al. 1993; Bullinger 1997). It is surprising that the quality of life approaches that are described have not been adopted in the area of health indicator research. From this

literature, three approaches for identifying relevant health items can be distinguished. In the sequential approach, an indicator that originates from one culture is used in another culture. In the parallel approach, each culture may contribute a specific indicator for a relevant health domain that is then adopted by all other countries. In the simultaneous approach, new health indicators are developed in a cross-cultural consensus.

Whichever approach is used, three basic steps are required in the development process: the translation of the indicator, its methodological testing and the norming process.

1.3.1 Translation

There has been extensive work on the issue of translation (Sartorius & Kuyken 1994; Guillemin et al. 1993). In cross-cultural sociological research, as well as in cross-cultural psychiatry and educational psychology, theoretical foundations and different methodological approaches to translating instruments from one culture to another have been developed (Brislin et al. 1973; Hui & Triandis 1985). Each working group has usually developed its own procedures for translation that are essentially based on a forward translation into the target language. However, the number of translators necessary as well as the use of back translations (from the target language back into the original language) is debated. In reviewing the different approaches to translation, Acquadro et al. (1996) emphasized the need to include at least two forward translations with a comparative discussion, and they were sceptical about the use of back translations, which are often hampered by potentially inferior quality that affects judgement of the forward translations. Most important is the international harmonization of translations in different countries by bringing together a group of bilingual persons from different countries who are able to interact and critically review each other's translations. While the basic philosophy of most guidelines for translation focuses on the adequacy of the translation from the original into the target language, Guyatt (1993) questioned the attempt to transpose a measure from one country to another as closely as possible, arguing that inconsistencies and illogical formulations as well as culturally non-transferable expressions can arise during translation, and suggested that these require an item to be reformulated (also in the original) rather than adapted into the target language. In spite of differences between translation approaches, most authors agree that the use of two forward translators is absolutely mandatory, the use of focus groups to evaluate the applicability of the translated health indicator in a specific country is recommended, and the use of back translation is debatable.

1.3.2 Methodological testing

The methods used to test the performance of survey instruments are based on approaches and procedures developed from psychometric theory. Such psychometric testing includes item descriptive statistics and scaling issues, as well as measures of reliability, validity and sensitivity (Bullinger et al. 1996). However, the various international working groups differ in the procedures they use for psychometric testing. In the area of quality of life, for example, the following approaches can be distinguished.

- The SF-36 Health Survey IQOLA Group (Aaronson et al. 1992) places importance on factor structure and item response theory as a means of distinguishing patterns of item responses across cultures (Ware et al. 1998). Emphasis is also placed on the performance of the questionnaires in terms of known group differences, i.e. testing

whether a measure is able to differentiate between patients who differ in degree of disease severity.

- The group that works with the FACT/FACIT questionnaire uses item analysis based on the Rasch model, structural equation models and multivariate statistics to replicate the factor structure of the measure across countries (Cella et al. 1996).
- The WHOQOL Project and also the SF-36 Group employ structural equation models to test the measurement model of the questionnaire across countries. In the WHOQOL work (WHOQOL Group 1994), that involves data from more than 4500 persons from 15 countries, the model is first fitted for the global data set and then replicated in each country (WHOQOL Group 1998). Item scale correlations and item descriptive statistics are also used to test whether items are applicable across cultures.
- In the EORTC Group, item and scale statistics have been used to decide whether specific items follow the measurement model in one country as compared to another (EORTC 1996).

1.3.3 Norming

A good example of cross-cultural norming studies is the SF-36 Health Survey. The IQOLA Group used population-based data to assess the quality of life of the general population in different countries. They found that the models that fitted western European data were very similar to one another. Intercountry comparisons of SF-36 sub-scores also showed a similarity in ratings, although with slight differences between the country profiles of industrialized western countries. The similarity of the SF-36 structure, as well as the convergent scale values across cultures, suggest that identical weighting systems can be used in these countries. The normative data can then be employed in each country to obtain age- and gender-specific reference data with which clinical quality of life data can be compared and expressed as deviations from the respective age- and gender-specific norms (Ware et al. 1996, 1998).

Comparison of quality of life data across countries as part of an epidemiological analysis is a challenge. This is different from comparative sociological research in which so-called “objective” indicators of quality of life are examined across countries. The availability of reference data makes it possible to examine, for example, whether differences in quality of life are due to socioeconomic conditions. Depending on the aim of the comparisons, such health determinants can be investigated to examine whether there is potential for intervention, or they can be statistically controlled for, in order to identify other factors that may be responsible for differences between cultures.

1.4 Prerequisites for health indicator harmonization and cross-population analysis

When developing cross-culturally applicable health indicators, it is important to review the underlying concept of health. Indicators are operationalizations of key components of health models. Such indicators can consist of a single statement or question, of several questions that can be aggregated into an index, or of a psychometric scale. An index is the accumulation of single statements, questions or observations into a more condensed form, using a derivation process that is based on a specific logic and criteria for the performance

of indicators (Bullinger 1992). A psychometric scale is a theoretically derived but empirically tested instrument that consists of items. It measures a construct and fulfils properties of scalability as well as reliability, validity and sensitivity. While the use of such scales is important in measuring health aspects that are not directly observable, the more easily observable indicators may also profit from the methodological rigor.

The major goals of health surveys are health monitoring and reporting, and thus health indicators are usually used descriptively in these contexts, each indicator being separately documented and analysed. However, it would be useful to supplement these descriptive goals with further analytical approaches that combine health indicators into more comprehensive health constructs. If the potential determinants of health were then added, a complex picture with a much richer analytical potential would emerge. Multidimensional health outcomes could be identified and linked to health determinants, thereby suggesting interventions that could be implemented and evaluated (Laaksonen et al. 2001). Furthermore, population subgroups (e.g. according to clinical condition, economic status, gender) could be studied for differences in health status. Cross-cultural analyses could also be performed, using each country individually, or grouping countries, or developing a pooled data set for all countries, where country is used as an explanatory factor in the analysis. Such statistical approaches could identify similarities and differences in health across countries, while at the same time taking into account interactions of health indicators and the influence of various other factors.

WHO has recently described specific methods for enhancing cross-population comparability of survey results in two documents (www.who.int/health-system-performance/technical-consultants/crosspopcomp_background.pdf and ... [draftreport.pdf](#)). These documents are based on the WHO Multi-Country Household Survey Study, which uses self-reported data for assessing population health status. An expert panel discussed ways in which the ordinal variables could be analysed (e.g. the ordered probit model from the social sciences, and the partial credit model from psychometrics). When these approaches were tested on a single data set, they yielded different results. A possible solution was the use of hierarchical ordered probit (HOPIT) and a hierarchical partial credit model (HPCM) on the basis of individual responses to vignettes. The discussions contrasted multi-item psychometric scales, in which differential item functioning (DIF) is a problem, with single-question approaches to health assessment, which would be necessary in surveys. Although these documents are highly relevant for health survey research, they do not yet contribute to the assessment of health as a complex construct. This could be achieved through the combination of indicators via latent class modelling. Sources of variation could then be detected and attributed to particular determinants. If the evolution of health indicators in a population were followed over time (e.g. with the use of longitudinal studies) it would also be possible to estimate the prognostic value of such determinants.

1.5 Discussion

There are a number of international efforts to assess cross-cultural health issues, but they are hampered by the fact that the majority of health surveys work with single indicators. Few of them use accumulated data (indices), and even fewer include psychometric instruments. Such instruments have been developed and are available in the quality of life and mental health fields. These instruments have passed the translation and initial testing phases and can now be reviewed for their cross-cultural performance. The first results, especially of the quality of life instruments, suggest good cross-cultural applicability.

While this chapter has focused on methodological issues that arise in the cross-population development and analysis of health indicators, there are also philosophical issues that need to be addressed. One such issue is the possible ethnocentrism of the health

indicator approaches used. The self-report questions that have been developed so far are mainly based on the verbal expressions of feelings or experiences from people in western countries, which could be perceived as the “middle-class Anglo-Saxon” outlook on health.

A second issue is the possible normativity of the health indicator concept. There is a concern that especially person-related data, for example on mental health, are not value-neutral but are standards to which the individual is expected to conform. If high function and wellbeing scores are equated with good health, then persons not fulfilling these standards could feel disadvantaged or discriminated against. In addition, biases in assessment may occur owing to the mode of questioning that is employed. For example, written questions may be not the most appropriate way to elicit health information – it may be that pictures or scenarios would be better alternatives.

A third issue concerns the ethical consequences of cross-cultural health research, where care needs to be taken in the interpretation of the health indicator information that is collected. Between-country comparisons have to be carefully modelled, taking into account differential effects of gender, age and socioeconomic conditions. Transparency of the methodology and the underlying concepts, as well as modesty, are mandatory in such health survey research and the researchers should also be prepared to take responsibility for their results after publication.

Health surveys have not only a descriptive function, but also a prescriptive aspect. As perceived by the individual, health is a human concern that is independent of race, gender, age, social status and occupation (Braveman et al. 2001). A moral imperative for societies is to scientifically study and politically address the health of its members. Not all societies are able to address the health of their citizens in a similar manner, especially countries in which living conditions need significant improvement. Health survey data can provide information about the respective status of populations and may indicate specific interventions by which health status can be improved. The effects of such interventions can then be measured and the resulting information can contribute to minimizing the gap between the “developed” and the “developing” world.

In conclusion, health is a universal human concept as far as its main dimensions are concerned, although the endorsement of these dimensions may differ across countries. This has implications for health policies in a globalizing world (Lee et al. 2002). Different societal conditions, cultural perspectives and types of individual behaviour may be relevant in different cultures, but these concern the ways rather than the aims of pursuing health. The achievement and maintenance of “good health” as a goal does not seem to vary culturally and is affected by nature and nurture as well as by the individual and the society. Health indicators should reflect as best as possible the essential components of health, critically examine their cross-cultural significance, and provide a sound base for health policy. The EUROHIS project, described in this book, attempts to contribute to these goals.

References

- AARONSON, N.K. ET AL. (1992) International quality of life assessment (IQOLA) project. *Quality of life research*, **1**: 349–351.
- ACQUADRO, C. ET AL. (1996) Language and translation issues. In: Spilker, B., ed. *Quality of life and pharmacoeconomics in clinical trials*, 2nd ed. Philadelphia, Lippincott-Raven, pp. 575–587.

- ANDERSON, R.T. ET AL. (1993) Critical review of the international assessment of health-related quality of life. *Quality of life research*, **2**: 369–395.
- AROMAA, A. (1998) Health observation and health reporting in Europe. *Revue d'épidémiologie et de santé publique*, **46**: 481–490.
- BORRELL, C. ET AL. (2000) Trends in social class inequalities in health status, health-related behaviors, and health services utilization in a Southern European urban area (1983–1994). *Preventive medicine*, **31**: 691–701.
- BRAVEMAN, P. ET AL. (2001) World Health Report 2000: how it removes equity from the agenda for public health monitoring and policy. *BMJ*, **323**: 678–681.
- BRISLIN, R.W. ET AL. (1973) *Cross-cultural research methods*. New York, John Wiley.
- BULLINGER, M. (1991) Quality of life – definition, conceptualization and implications – a methodologist's view. *Theoretical surgery*, **6**: 143–148.
- BULLINGER, M. (1992) Indices and profiles: advantages versus disadvantages. In: Walker, S., ed. *Key issues in quality of life assessment*. Dordrecht, Kluwer Academic Publishers, pp. 215–226.
- BULLINGER, M. (1997) The challenge of cross-cultural quality of life assessment. *Psychology and health*, **12**: 815–825.
- BULLINGER, M. ET AL. (1996) Creating and evaluating cross-cultural instruments. In: Spilker, B., ed. *Quality of life and pharmacoeconomics in clinical trials*, 2nd ed. Philadelphia, Lippincott-Raven, pp. 659–669.
- CAVELAARS, A.E. ET AL. (1998) Differences in self reported morbidity by educational level: a comparison of 11 western European countries. *Journal of epidemiology and community health*, **52**: 219–227.
- CELLA, D.F. ET AL. (1996) Cross-cultural instrument equating current research and future directions. In: Spilker, B., ed. *Quality of life and pharmacoeconomics in clinical trials*, 2nd ed. Philadelphia, Lippincott-Raven, pp. 707–716.
- CORNIA, G.A. (2001) Globalization and health: results and options. *Bulletin of the World Health Organization*, **79**: 834–841.
- CURRIE, C. ET AL., ED. (2001) *Health and health behaviour among young people*. Copenhagen, WHO Regional Office for Europe (Health Policy for Children and Adolescents Series No. 1; document EUR/ICP/IVST 06 03 05(A)).
- DONNAN, S. (1998) The health of adult Europe. Combating inequalities involves measuring what counts. *BMJ*, **316**: 1620–1621.
- ELOLA, J. ET AL. (1995) Health indicators and the organization of health care systems in Western Europe. *American journal of public health*, **85**: 1397–1401.
- EUROPEAN ORGANISATION FOR RESEARCH AND TREATMENT OF CANCER (EORTC) (1996) Modular approach to quality of life assessment in oncology: an update. In: Spilker, B., ed. *Quality of life and pharmacoeconomics in clinical trials*, 2nd ed. Philadelphia, Lippincott-Raven, pp. 179–191.
- GUDEX, C. & LAFORTUNE, G. (2000) *An inventory of health and disability-related surveys in OECD countries*. Paris, Organisation for Economic Co-operation and Development (OECD Labour Market and Social Policy, Occasional Paper No. 44, DEELSA/ELSA/WD(2000)5).
- GUILLEMIN, F. ET AL. (1993) Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of clinical epidemiology*, **46**: 1417–1432.
- GUYATT, G.H. (1993) The philosophy of health-related quality of life translation. *Quality of life research*, **2**: 461–465.
- HUI, C. & TRIANDIS, H.C. (1985) Measurement in cross-cultural psychology: a review and comparison of strategies. *Cross-cultural psychology*, **16**: 131–152.
- HUPKENS, C.L. ET AL. (1999) National health interview surveys in Europe: an overview. *Health policy*, **47**: 145–168.
- HUTCHINSON, J.E. (1996) Quality of life in ethnic groups. In: Spilker, B., ed. *Quality of life and pharmacoeconomics in clinical trials*, 2nd ed. Philadelphia, Lippincott-Raven, pp. 587–595.
- JOHNSON, T.M. (1996) Cultural considerations. In: Spilker, B., ed. *Quality of life and pharmacoeconomics in clinical trials*, 2nd ed. Philadelphia, Lippincott-Raven, pp. 511–517.
- LAAKSONEN, M. ET AL. (2001) Do health behaviour and psychosocial risk factors explain the European east-west gap in health status? *European journal of public health*, **11**: 65–73.
- LEE, K. ET AL., ED. (2002) *Health policy in a globalising world*. Cambridge, Cambridge University Press.
- MURRAY, J.L. & LOPEZ, A.D. (1995) The global burden of disease: a comprehensive assessment of mortality and disability from disease, injuries and risk factors, 1990 and projected to 2020. Geneva, World Health Organization.
- ORLEY, J. & KUYKEN, W., ED. (1994) *Quality of life assessment: international perspectives*. Berlin & Heidelberg, Springer.

- OVRETVEIT, J. (2001) Quality evaluation and indicator comparison in health care. *International journal of health planning and management*, **16**: 229–241.
- RAVENS-SIEBERER, U. ET AL. (2001) Quality of life in children and adolescents: a European public health perspective. *Sozial- und Präventivmedizin*, **46**: 294–302.
- SARTORIUS, N. & KUYKEN, W. (1994) Translation of health status instruments. In: Orley, J. & Kuyken, W., ed. *Quality of life assessment: international perspectives*. Berlin & Heidelberg, Springer, pp. 3–19.
- STEWART, A. & WARE, J. (1992) *Measuring function and wellbeing*. Durham, NC, Duke University Press.
- WARE, J.E. JR ET AL. (1996) Evaluating instruments used cross-nationally: methods from the IQOLA Project. In: Spilker, B., ed. *Quality of life and pharmacoeconomics in clinical trials*, 2nd ed. Philadelphia, Lippincott-Raven, pp. 681–693.
- WARE, J.E. JR ET AL. (1998) The factor structure of the SF-36 Health Survey in 10 countries: results from the IQOLA Project. *Journal of clinical epidemiology*, **51**: 1159–1165.
- WHOQOL GROUP (1994) The development of the World Health Organization Quality of Life Assessment Instrument (the WHOQOL). In: Orley, J. & Kuyken, W., ed. *Quality of life assessment: international perspectives*. Berlin & Heidelberg, Springer, pp. 41–61.
- WHOQOL GROUP (1998) The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. *Social science and medicine*, **46**: 1569–1585.
- WORLD HEALTH ORGANIZATION (1985) *Constitution of the World Health Organization*. Geneva, World Health Organization, pp.1–2.
- WORLD HEALTH ORGANIZATION (2000) *The world health report 2000 – health systems: improving performance*. Geneva World Health Organization.
- ZWI, A.B. & YACH, D. (2002) International health in the 21st century: trends and challenges. *Social science and medicine*, **54**: 1615–1620.

Chapter 2.

Overview of the EUROHIS project

Anatoliy Nosikov and Claire Gudex
World Health Organization Regional Office for Europe, Copenhagen, Denmark

Health interview surveys remain a unique source of data for some essential aspects or “indicators” of health, such as health behaviour and attitudes, patterns of use of health services and consumption of health-influencing substances. (The term “indicator” is used here to represent the entire area under a broad concept: this includes the definition of the concept and approaches to its measurement, as well as methods of collecting, processing and presenting data on that concept.) Surveys are a major source of information for governments to fulfil their stewardship role in identifying policy priorities, and in planning, monitoring and evaluating policy responses. But even the most sophisticated national information systems are not sufficient in themselves to support this function. All agree that international comparability of health data and indicators is necessary. From a European health policy perspective, at least three issues should be emphasized.

First, the availability of internationally comparable health data can highlight true variations in health status and the inputs and factors associated with different health outcomes across countries. Such information can then guide national policy-related discussions on best practices to achieve the maximum possible output and to improve the health of populations.

Second, the lack of reliable comparative findings hinders the harmonization of European policies and strategies, and the monitoring of measurable priority objectives to track progress and adjust strategies. This can also severely reduce the credibility of currently available recommendations for action, and thus weaken the support for policy responses.

Third, the need to improve international comparability is only the more visible part of the need to improve cross-cultural comparability, particularly within countries. The problems related to comparability of data between social groups are generally underestimated. This issue is important in any endeavour to change health behaviour, where multifaceted interventions must precisely target cultural and demographic differences.

Despite the advantages of collecting internationally comparable data, European national population surveys have tended to use different instruments to measure the same indicators, different survey modes and methodologies, and different reporting formats. Instruments have frequently been developed ad hoc for the specific context, or have been adapted from other countries without international coordination. This has made international comparison of European national survey results difficult or even impossible. It may well benefit Europe to change its focus away from trying to compare a variety of national health surveys and towards the application of international multi-country surveys. For a number of indicators,

this may be a better approach (technically, managerially and economically) to achieving international comparability.

Even if the political mandate and financial resources for such a regular multi-country survey are secured, national health interview surveys will of course continue to be commissioned, as national policy-makers will still need more detailed information on specific elements of their own health system. This will be especially important in the face of the complex social changes that are under way in many European countries. Under these circumstances, a national health survey can provide a more flexible monitoring tool. However, it is certain that as common European policies and strategies evolve, increasing weight will be given to common survey instruments, either as modules in national health surveys or as part of Europe-wide surveys, as well as to the issues of cross-population comparability (see, for example, Murray et al. 2002a, 2002b).

2.1 Background to the EUROHIS project

The adoption by WHO in 1977 of the strategy for health for all highlighted the need for internationally comparable health data that were derived from population surveys. To facilitate the development and use of common methods and instruments for health interview surveys, a series of three consultations was organized in 1988, 1990 and 1992 by Statistics Netherlands under the aegis of the WHO Regional Office for Europe. Together with the subsequent follow-up, these consultations produced specific recommendations regarding common instruments for 11 health for all indicators. The next step in this process was to promote the actual implementation of these instruments in the national health interview surveys of WHO's European Member States, and to evaluate their performance. It was realized that such an evaluation could further promote the use of common instruments, and would enhance the comparability of survey data in general.

An investigation in 1992 by the Regional Office indicated that, for about half of the 11 health for all indicators, no or only relatively minor measurement differences existed between countries. On this basis a further study was initiated to ascertain how comparable the data on these indicators were, what differences in measurement still existed, and what the possibilities were for achieving widespread use of the recommended instruments. This study was conducted by Statistics Netherlands together with the other countries participating in the consultations and the Regional Office. Six of the 11 health for all indicators were reviewed: perceived health, temporary disability, smoking, birth weight, breastfeeding and body mass index. A questionnaire was sent to the countries concerned at the end of 1993, along with the data request for the 1993/1994 health for all monitoring process. The following information was requested for each of the six indicators:

- € details of the most recent survey for which results on the indicator were available;
- € a copy of the full questionnaire and the exact wording of the instrument in English;
- € the mode of data collection (face-to-face interview, telephone interview, self-administered questionnaire or other);
- € target population and sample size;
- € percentage of non-response;
- € percentage of proxy interviews; and
- € the most recent outcome data by age and gender.

This evaluative study provided specific information on the comparability of survey data on these six indicators, including some apparently comparable data sets, and thus encouraged the future use of similar approaches for acquiring internationally comparable data. The data collation exercise was repeated in 1995/1996.

In response to the need for a review of the objectives and plan of work in the harmonization of methods and instruments, a review of the achievements made so far were published in the “blue book” (de Bruin et al. 1996). The experience accumulated by the project at that time naturally led to the desire to expand the harmonization process to include other key indicators that had not been included in the previous rounds because of their greater complexity. An important issue was the selection of further indicators for harmonization. Thus, in 1995, an inventory of surveys was conducted to determine the opinions of the participants with regard to future priorities. Replies were received from 25 of the 31 countries approached. All respondents indicated interest in the project, and all but one agreed to supply data for international comparison.

In 1997, a fourth meeting in the series of consultations was held to review progress to date and to discuss the possibility of developing common instruments for the remaining health for all indicators. The results of the 1995 inventory study were useful in determining a plan for the next stage of work in harmonizing European health interview surveys (HIS, hence the new name of the EUROHIS project). It was agreed that, under the principal objective of developing and promoting the use of common instruments in Europe, the two sub-objectives should be to:

- € develop recommended common instruments for national health interview surveys, and
- € develop methods to adjust national data to allow cross-national comparisons.

The project should be divided into stages that would reflect the tasks under each of the two sub-objectives. Thus, the development of recommended common instruments should comprise, for each indicator:

- € identification of a principal investigator and the establishment of a research group;
- € a review of the instruments that were currently being used in Europe;
- € selection and field-testing of a common instrument; and
- € publication of the recommended common instrument.

The second sub-objective, the adjustment of national data, should comprise, for each indicator:

- € finalization of instructions for data collection and adjustment;
- € collection of data from countries by mail;
- € processing and adjustment of data for international data sets; and
- € a review of the results and publication of the comparable data.

This plan formed the basis for a subsequent EC/DGXII/BIOMED 2 proposal, which was accepted for funding as a concerted action over three years, starting in 1998. Common instruments for the following eight indicators were to be developed: chronic physical conditions, mental health, alcohol consumption, physical activity, use of curative medical services, use of medicines, use of preventive health care and quality of life.

2.2 The EUROHIS approach

It was recognized that, while the first round of international consultations had relied mainly on focus group discussions to remove unnecessary differences between methods and instruments, the next round should delve more deeply into the conceptual and health system backgrounds against which the individual countries developed and validated their instruments. The task of defining the conceptual content for the topic under measurement is even more important in a comparative cross-national study than in single-nation research. Projects such as EUROHIS represent a specific type of cross-national survey with a focused integrated design, in which groups of collaborators from different countries make decisions jointly and all participants follow similar steps at nearly the same time (Glaser 1977). This kind of collaboration is fundamental in a cross-national research effort that aims to develop common instruments for use in the participating countries.

In view of the size and complexity of the project, the management structure comprised three layers. First, a scientific and management board composed of the principal investigators for each indicator and special project advisers. Second, a network group for each indicator composed of the respective principal investigator and a number of active participants. Third, individual organizations in 33 European Member States who provided regular feedback on progress and also participated in testing the performance of the instruments.

2.2.1 Stage A of the EUROHIS project

Stage A (the development of recommended common instruments) was undertaken between June 1998 and March 2001. It included the following steps:

- € concept exploration and working definition of each indicator;
- € review of instruments currently used in Europe (the “survey of surveys”);
- € development of draft instruments;
- € pre-testing in 18 European countries¹;
- € international meeting to review the progress made (mid-term review); and
- € production of revised draft instruments.

These main steps formed the basic programme of work for each of the indicator networks. As will be seen in the following chapters, however, the method of developing the instruments has varied, owing either to the nature of the indicator itself or to the opportunity to build on experiences from other projects. Thus the indicator network groups for mental health and quality of life have worked on instruments that were already established, while the instruments for alcohol consumption and physical activity have been developed alongside other ongoing international efforts. In contrast, the instruments for chronic physical conditions, use of curative medical services, use of medicines and use of preventive health care have been developed solely as part of the EUROHIS project.

Despite the different paths taken by the indicator network groups, there were common stages of work for all the instruments. These are described here to avoid repetition in the later chapters related to each indicator.

¹ ‘European’ in this context and throughout the rest of the book refers to countries that are Member States of the WHO European Region (for list of Member States of the WHO European Region please refer to <http://www.euro.who.int/AboutWHO/About/MH>)

The survey of surveys

This review of existing instruments used in national health interview surveys was carried out in 1999/2000 in collaboration with the EC-project “Health surveys in the EU: HIS and HIS/HES evaluations and models” and the inventory of surveys of Statistics Netherlands. This limited the reporting burden on the Member States and broadened the scope of information that was obtained. The main output was a comprehensive inventory of 30 national surveys from 25 countries, and feedback from the survey managers in the Member States about their requirements and preferences for the instruments being developed by EUROHIS.

Pre-testing of draft instruments in 2000

The draft instruments were first tested in 2000 on what were mainly small convenience samples in 18 countries: Bosnia and Herzegovina, Bulgaria, the Czech Republic, Estonia, Finland, Germany, Hungary, Italy, Latvia, Lithuania, the Netherlands, Romania, the Russian Federation, Slovakia, Slovenia, Spain, the former Yugoslav Republic of Macedonia and Ukraine. A variety of methods were used and there were differences between countries in the place and mode of interviewing, the way of recruiting respondents, and the level and structure of follow-up or check questions. The number of respondents ranged from 10 in the Russian Federation to 133 in Bulgaria; most countries used about 20, which proved to be sufficient to identify the main issues. As much as possible, the instruments were tested as a group; different numbers of countries, however, tested different combinations of instruments (Table 2.1). The pre-test proved extremely useful in identifying sections or phrases in the instruments that were ambiguous, difficult or inappropriate to translate into other languages or cultures, or required changes in question order or routing.

Table 2.1 Pre-testing of draft EUROHIS instruments, 2000

Number of countries	Draft instruments tested^a
10	1. Chronic physical conditions 5. Use of curative medical services 6. Use of medicines 7. Use of preventive health care 8. Health-related quality of life
5	1. Chronic physical conditions 4. Physical activity 5. Use of curative medical services 6. Use of medicines 7. Use of preventive health care 8. Health-related quality of life
1	1. Chronic physical conditions 5. Use of curative medical services 7. Use of preventive health care 8. Health-related quality of life
1	1. Chronic physical conditions 7. Use of preventive health care 8. Health-related quality of life

^a: In addition, Germany tested one indicator, that of use of medicines.

Mid-term review

The results from the EUROHIS pre-testing were presented and discussed at this meeting, which was held on 2–4 November 2000 in Copenhagen. The meeting was attended by participants from 27 countries in the WHO European Region, and included the project active participants and principal investigators, the survey managers from CEE and NIS countries who had pre-tested the draft EUROHIS instruments, and staff from WHO and the Organisation for Economic Co-operation and Development (OECD).

2.2.2. Stage B of the EUROHIS project

Stage B was initially planned as the collation and adjustment of existing data on the indicators to allow international comparisons. During the course of Stage A, however, it became clear that these goals would be extremely difficult to fulfil. The draft EUROHIS instruments contained many new elements on which data had not previously been collected. In addition, existing country data were too variable in content and quality to enable the harmonization exercise to yield results that would justify the effort and difficulty of the task. Discussions at the mid-term review in November 2000 led to three options for stage B of the project: extending the survey of surveys, undertaking more extensive field-testing of the instruments and/or encouraging implementation of the draft instruments in national health surveys. While these options were not mutually exclusive, it was decided that emphasis should be placed on more extensive field-testing of the draft instruments, with the aim of obtaining empirical evidence of their performance in the field.

Field-testing of draft instruments in 2001

Standard protocols for both CATI (computer-assisted telephone interview) and CAPI (computer-assisted personal interview) approaches were finalized in May 2001. These protocols described methodology for translation, sample selection and data collection (including observations of both interviewer and respondent behaviour). A data coding book and a data analysis plan were developed for use by all field-testers.

The eight EUROHIS indicators were combined into a single questionnaire in English. To minimize the burden on respondents, the optional items of some of the EUROHIS indicators were omitted, resulting in a questionnaire that required about 30 minutes in a telephone interview. Before translation, the questionnaire was edited to check grammar and appropriate wording. The aim of the translation process was to achieve conceptual equivalence in each of the target cultures, i.e. a common understanding of the nature and content of the concepts that were measured by the instrument (Herdman et al. 1998). This was a natural continuation of the EUROHIS instrument development process, which had involved researchers from a range of cultures as well as the examination of national health surveys currently used across Europe. The translation process involved four main steps: (a) forward translation into the target language, by a native speaker of that language; (b) back translation into English, by a native English speaker with no previous knowledge of the questionnaire; (c) review by Regional Office staff of back translations; and (d) production of the final version in the target language. Back translation proved to be extremely useful in clarifying the interpretation of many words and terms, as well as increasing the similarity between the different language versions.

Field-testing was undertaken in 15 European countries. The internationally operating market research institute, NIPO, field-tested in France, Germany and the United Kingdom in November and December 2001, using a standardized CATI approach and a sample size of 400 completed interviews in each country (randomly selected nationally

representative samples; age range 35–75 years). Eleven CCEE/NIS countries (Azerbaijan, Croatia, Czech Republic, Kyrgyzstan, Latvia, Lithuania, Republic of Moldova, Romania, Russian Federation, Slovakia and the former Yugoslav Republic of Macedonia) field-tested the instruments between July and October 2001 using mainly face-to-face interviews (Lithuania conducted a postal survey) and either convenience samples or national random samples of the adult population. The number of completed questionnaires ranged from 287 to 656. Israel field-tested in November/December 2001 and used a CATI approach for 924 interviews.

The field-test data were collected as a central database, with analysis being undertaken centrally as well as by the field-testers and by the principal investigators for each instrument. It should be noted that only the surveys that were conducted in France, Germany and the United Kingdom used the same methodology. Although an attempt was made to harmonize the methods used in the EUROHIS field-test, the other surveys conducted had different sample sizes and strategies, used different modes of administration and produced either weighted or unweighted data (see section 4.5.1 and Table 4.1 for further details). The results from all the field-test countries have of course been useful in assessing the performance of the EUROHIS instruments, but reliable comparisons of the actual indicator data can only be made using the data from France, Germany and the United Kingdom. In the chapters that follow, therefore, conclusions from the field-test are mainly based on the data from these three countries.

2.3 Output from the EUROHIS project

The considerable amount of material produced over the course of the EUROHIS project is freely available for use by others. Apart from the present book, this material includes

- € a wide-ranging review of existing methods and instruments used in national surveys to measure seven main health indicators (the so-called “survey of surveys”);
- € official WHO translations into French, German and Russian of the “long” field-testing questionnaire (i.e. all items);
- € translations of the “short” field-testing questionnaire (i.e. optional items excluded) into Arabic (Israel), Azerbaijani, Croatian, Czech, Hebrew (Israel), Kyrgyz, Latvian, Lithuanian, Macedonian, Romanian (field-tested in the Republic of Moldova and Romania), Russian (field-tested in the Republic of Moldova and the Russian Federation), Slovak and Ukrainian;
- € national reports on field-testing in each country in 2001;
- € 2001 field-test data (note that publications based on these data require appropriate reference to the EUROHIS project and acknowledgement of financial contribution from the European Commission);
- € The CATI and CAPI protocols, coding book and analytical framework used in the field-testing;
- € reports and data from the 2000 pre-test of the draft EUROHIS instruments;
- € minutes of the six Board meetings held during the course of the project; and
- € the two annual reports and the final project report produced for the European Commission.

Further details of this material and of the EUROHIS project are available on request from the WHO Regional Office for Europe.

References

- DE BRUIN, A. ET AL. ED. (1996) *Health interview surveys: towards international harmonization of methods and instruments*. Copenhagen, WHO Regional Office for Europe (WHO Regional Publications, European Series, No. 58).
- GLASER, W.A. (1977) The process of cross-national survey research. In: Szalai, A. & Petrella, R., ed. *Cross-national comparative survey research: theory and practice*. Oxford, Pergamon Press, pp. 403–435.
- HERDMAN, M. ET AL. (1998) A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Quality of life research*, 7: 323–335.
- MURRAY, C.J.L. ET AL. (2002a) *Enhancing cross-population comparability of survey results*. Geneva, World Health Organization (GPE Health Policy Discussion Paper No. 35; http://www3.who.int/whosis/discussion_papers/pdf/paper35.pdf, accessed 1 October 2002).
- MURRAY, C.J.L. ET AL. (2002b) *Cross-population comparability of evidence for health policy*. Geneva, World Health Organization (GPE Health Policy Discussion Paper No. 46; http://www3.who.int/whosis/discussion_papers/pdf/paper46.pdf, accessed 1 October 2002).

Chapter 3.

Development of a common instrument for chronic physical conditions

Vittoria Buratta, Luisa Frova, Lidia Gargiulo,
Emilio Gianicolo, Sabrina Prati and Luciana Quattrociochi
Italian National Institute of Statistics (ISTAT), Rome, Italy

3.1 Introduction

Chronic disease represents one of the most relevant problems for health-related quality of life, especially for the elderly, and one of the main reasons for the use of health services. Chronic disease often affects the psychological or motor skills of people for a long time or even indefinitely. Although it is important to monitor the distribution and trends of chronic disease in the general population, few countries collect such morbidity data on a regular basis. Data are usually collected only for selected diseases and for subgroups that are not necessarily representative of the whole population (for example, for the purposes of cancer registers).

Many countries collect data on morbidity from chronic physical conditions through the use of interview surveys. This approach is particularly suitable for diseases of long duration and low mortality and is also cheap and relatively quick to implement, and allows the simultaneous collection of other health-related information such as use of medical services. Interview surveys also have some disadvantages, however. The sample method does not permit measurement of conditions with a low prevalence, and the subjective reporting may not always correctly identify the condition, for example people may have a disease but be not aware of it, or they may have incorrectly diagnosed a condition themselves. Furthermore, there is considerable heterogeneity in the concepts and definitions that are adopted in different countries when surveying chronic conditions.

The objective of the EUROHIS network on “chronic physical conditions” was to develop and recommend a common instrument to measure this indicator in health interview surveys (HIS).

3.2 Methodology

The work on this topic was carried out in several phases. First, under concept exploration, it was discussed what was meant by a “chronic condition”. This led to a discussion of the key methodological aspects, such as the correct methodological approach for surveying chronic conditions, the most appropriate reference period and the criteria for reporting. The so-called “survey of surveys” provided a review of instruments currently used in European countries to measure chronic physical conditions, and formed the basis for comparisons between the questions and methods used in different countries. A preliminary common instrument to survey chronic physical conditions was then developed and tested, first in a pre-test in 2000 and then in a larger field-test in 2001. The performance of the instrument

was assessed in different national environments, and a final recommended common instrument to survey chronic physical conditions was proposed.

3.3 The survey of surveys

A mail survey was conducted among the European countries. Information was collected on the definitions used in surveying chronic conditions, other data sources on chronic conditions available in each country, and comparability between HIS and other sources. Nine countries responded with information on chronic conditions: Austria, Belgium, Finland, Germany, Italy, the Netherlands, Spain (Catalonia), Switzerland and Ukraine.

The responses confirmed that there is considerable variation in the definitions used for chronic conditions. Very few countries reported explicit criteria, usually referring to the nature of the disease and its duration. Nevertheless, there was general agreement that chronic conditions are of long duration or permanent, and are expected to require a long period of supervision, observation or care. Most of the countries sent a copy of the instrument that they used to survey chronic conditions.

All the countries reported that data on chronic conditions were available from hospital statistics and most of them also had register data. One third of the countries also indicated other sources of data, such as epidemiological studies or ad hoc surveys.

Fewer than half the countries had conducted comparative studies on their data from different sources. In general, these data were not directly comparable, owing to the different methodology used to collect them. From the information provided by Finland, Italy and the Netherlands, it appeared that registers and interview surveys have their own advantages and limitations, and the level of comparability depends on the nature and prognosis of the disease (Table 3.1).

Table 3.1. Comparability between HIS and other data sources in Finland, Italy and the Netherlands

Country	Comparisons	Outcome
Finland	HIS vs. hospital inpatient care	The register and survey data are not directly comparable, but both data sets are valuable. Analyses of the socioeconomic differences in both data sets show similar trends.
	HIS vs. HES ^a	Overall prevalence of chronic morbidity is the same whether measured through self-reporting or by doctors' examinations, but discrepancies arise at the individual level.
	HIS vs. register of social insurance institutions or hospital discharge register	Self-reporting is valid for many clearly defined diseases (e.g. diabetes) but differences occur for conditions such as musculoskeletal disease and mental disorders (Heliövaara et al. 1993).

Italy	HIS vs. HES (Maggi et al. 1994)	The level of agreement differs by disease. Large differences in prevalence are found for hypertension (probably because this condition is asymptomatic for most people), smaller differences for angina pectoris and perfect concordance for diabetes. Proxy responses in HIS typically produce lower prevalence of less severe conditions (e.g. allergy, arthrosis).
Netherlands	Health survey of Statistics Netherlands vs. general practitioner register	The level of agreement differs by disease (or disease group) and is affected by the nature and prognosis of the disease (e.g. good agreement for diabetes, epilepsy and cancer, but less agreement for respiratory disorders, hypertension and migraine). Both registers and health surveys have advantages and limitations. The different observation periods are a weak point in the comparisons.

^a Health examination survey.

3.4 Further review of instruments recently used in European studies

A further international comparison was made by using the survey of surveys data to update an earlier EUROSTAT study that had compared measurement methods used in European health surveys (Hupkens 1998). This comparison involved 18 HIS surveys from 17 countries: Austria 1999, Belgium 1997, Denmark 1994, England 1996, Finland 1996 (two surveys), France 1998, Germany 1998, Ireland 1998, Italy 1999/2000, the Netherlands 1999, Norway 1995, Portugal 1995/1996, Scotland 1996, Spain 1997, Sweden 1996, Switzerland 1997 and Ukraine 1999.

3.4.1 Types of question

In surveying chronic conditions, two types of question are used: open-ended questions and disease-specific questions. The open-ended question enquires whether respondents have chronic conditions and, if they do, what these are. The disease-specific question comprises a check-list of conditions or diseases (or a card that has to be shown) and respondents are asked whether or not they have one or more of them.

Of the 18 survey questionnaires that were reviewed:

- seven surveys included only a disease-specific question;
- four surveys included only an open-ended question, and the respondents were further asked to define and specify each condition they had; and
- seven surveys included both open-ended and disease-specific questions.

In the last seven surveys all the respondents, whether or not they had declared a chronic condition, had to also answer the disease-specific questions (i.e. the general open-ended question did not act as a filter for the disease-specific questions).

The open-ended questions that were used by the different countries rarely referred explicitly to physical and/or mental conditions, and instead asked generally about any “chronic condition, defect or injury”. In most cases, the open-ended questions explicitly

included the terms “handicap” or “disability”.

In the disease-specific questions, the check-lists included different types of conditions, including physical conditions, mental problems and disabilities. The terminology varied widely between countries, with check-lists including clinical diagnostic terms and/or symptoms and/or complaints.

3.4.2 Criteria for reporting diseases

Ten surveys included a question on diagnosis of chronic diseases. In three of these, the disease-specific question was restricted to diseases that had been diagnosed by a doctor. In six surveys, respondents were asked whether or not they had received treatment for any specific conditions. Only one survey inquired about hospital admission.

3.4.3 Reference period

The open-ended questions usually referred to the present time, with the exception of one survey (which referred to the past two weeks). The disease-specific questions referred to both present and past states of health in five surveys, to the past 12 months in six surveys, to the past six months in one survey, and to the present time only in two surveys.

3.4.4 Administration methods

Most of the surveys were conducted through face-to-face interviews. Often a combined method of administration was used, i.e. face-to-face and self-administered in five surveys, face-to-face and telephone in one survey, self-administered and telephone in one survey, and all three administration methods in France. Half of the surveys allowed proxy answers for children, and often also for adults who were unable to answer or who were not at home.

3.4.5 Other methodological aspects

All 18 HIS were cross-sectional. Four countries combined a cross-sectional survey with a follow-up. The type of sample was a household sample in nine surveys and an individual sample in the other nine surveys.

3.4.6 Disease check-lists

Fifteen surveys included a check-list of diseases, but these differed with respect to the number of diseases included and the type of disease. The number of diseases varied between 7 and 46. More than half of the surveys included more than 20 diseases. Table 3.2 shows the major groups of diseases that were included in the 15 check-lists examined and the diseases that were investigated in each group.

An analysis of the disease groups that were most frequently investigated showed general agreement on the definitions and criteria used for chronic diseases. At least one of the following criteria could be found: prevalence, severity, economic cost, use of health care services and amenability to self-reporting.

All of the surveys included at least one cardiovascular disease. Hypertension was

investigated in all surveys. Questions about ischaemic heart disease (heart attack, angina) and stroke were frequently found (11, 10 and 7 times, respectively). Usually 20% (or more) of the specific diseases included in the check-list referred to the circulatory system. The terminology that was used varied among countries and could include formal medical terms or descriptions of symptoms and/or complaints. In eight surveys, respondents were able to add diseases that were not covered by the check-list.

Table 3.2. Diseases and condition groups that were included in 15 European surveys, and the specific diseases most frequently listed

Disease or condition group	No. of surveys	Specific disease cited and No. of surveys
Cardiovascular diseases	15	Hypertension, ischaemic heart disease (angina, heart attack), stroke
Metabolic disorders	13	Diabetes (13), thyroid trouble (5)
Respiratory diseases	13	Bronchial asthma (10), chronic bronchitis (11)
Pain and other musculoskeletal troubles	12	Back pain (11), arthritis/arthrosis (7), osteoporosis
Mental disorders	10	Depression/anxiety (9)
Diseases of the nervous system	10	Migraine (9), epilepsy (7)
Diseases of the digestive system	10	Gastric or duodenal ulcer (8), gallstones
Diseases of the urogenital system	10	Kidney stones (10), chronic renal diseases (glomerulonephritis, nephropathy, pyelonephritis, renal insufficiency), prostate
Allergies	9	Hay fever, contact dermatitis, food allergies, etc.
Cancer	8	Cancer
Skin diseases	7	Eczema, psoriasis, bullous dermatosis, vitiligo, skin ulcer, etc.
Other conditions, handicaps	6	Sight troubles, hearing troubles, amputations, etc.

3.5 Preliminary common instrument for chronic physical conditions

The indicator network for chronic physical conditions agreed on the inclusion of two questions in the preliminary common instrument.

Open-ended question

Do you have any long-standing illness or health problem? No, Yes

For those countries that use “chronic” as a more general term than “long-standing”, the term “chronic” was recommended, for example France and Italy.

Disease-specific question

Do you have or have you ever had any of the following chronic illnesses or conditions?

(reply to each of the illnesses)

Diseases or conditions	Do you have or have you ever had any of these following conditions?		If YES, then:					
			Was it diagnosed by a physician?		Do you take drugs or are you under therapy?		Did you have this problem in the past 12 months?	
1. Diabetes	No	Yes	No	Yes	No	Yes	No	Yes
2. Hypertension (high blood pressure)	No	Yes	No	Yes	No	Yes	No	Yes
3. Heart attack (myocardial infarction)	No	Yes	No	Yes	No	Yes	No	Yes
4. Stroke (cerebral stroke, cerebral haemorrhage and other cerebrovascular diseases)	No	Yes	No	Yes	No	Yes	No	Yes
5. Malignant tumour/cancer (including leukaemia and lymphoma)	No	Yes	No	Yes	No	Yes	No	Yes
6. Asthma (including allergic asthma)	No	Yes	No	Yes	No	Yes	No	Yes
7. Chronic bronchitis, emphysema	No	Yes	No	Yes	No	Yes	No	Yes
8. Arthrosis, arthritis	No	Yes	No	Yes	No	Yes	No	Yes
9. Chronic anxiety and depression	No	Yes	No	Yes	No	Yes	No	Yes
10. Migraine and frequent headache	No	Yes	No	Yes	No	Yes	No	Yes
11. Serious skin diseases	No	Yes	No	Yes	No	Yes	No	Yes
12. Allergy (including allergic asthma)	No	Yes	No	Yes	No	Yes	No	Yes
13. Osteoporosis	No	Yes	No	Yes	No	Yes	No	Yes
14. Cataract	No	Yes	No	Yes	No	Yes	No	Yes
15. Gastric or duodenal ulcer	No	Yes	No	Yes	No	Yes	No	Yes
16. Other (specify)	No	Yes	No	Yes	No	Yes	No	Yes
.....								
.....								
.....								

3.6 EUROHIS pre-test 2000

The main purpose of the pre-test was to explore any difficulties that the respondents had in understanding or answering the questions in the draft instrument. It was possible that problems would arise as a result of translating English terms into the local language or because of ambiguity in some of the questions proposed. The draft instrument for chronic physical diseases was combined with several of the other draft EUROHIS instruments and tested in 17 European countries (a detailed description of the Italian experience is presented in section 3.6.1).

The results did not indicate any translation problems for the preliminary instrument on chronic conditions, and there was general agreement among countries on the appropriateness of the questions proposed. Most of the countries appreciated the robustness and the practicability of the instrument as a result of its brevity and clarity.

Some issues did arise, however, regarding the disease-specific question.

- The order in which the diseases were listed could be improved. Some serious chronic diseases such as cancer were listed at the beginning of the list; at times this caused negative feelings and the respondent then did not provide information on the other chronic diseases listed thereafter. In response to this issue, a revised ordering of the diseases in the check-list was tested in Italy (see section 3.6.1 and Table 3.5).
- Items 6 “Asthma (including allergic asthma)” and 12 “Allergy (including allergic asthma)” overlapped and this could produce problems. As a very high percentage of asthma is “allergic asthma”, item 6 could have an additional question “If yes, is it allergic asthma?” Item 12 should then be “Allergy (excluding allergic asthma)”. This option was also tested in Italy.
- Many countries found the check-list too short and suggested the addition of further diseases such as chronic kidney and liver diseases, Parkinson’s disease and epilepsy. It was agreed that countries could add other diseases but should put them at the end of the check-list before “other” diseases.

A further issue arose about the question “Do you take drugs or are you under therapy?” It was considered that this question could overlap with some of the questions included in two of the other EUROHIS indicators: use of curative services and use of medicines. However, the aim of this question was to explore severity of disease and not to measure the use of medicine or therapy. The question could be divided into two different questions (i.e. “Do you take drugs?” and “Are you under therapy?”) or indeed could be excluded completely as it is an optional item.

3.6.1 Analysis of the Italian pre-test results

The Italian pre-test was conducted between 16 and 20 October 2000, with the aim of assessing the validity of the draft instruments for five EUROHIS indicators, as outlined above.¹ Personal interviews were conducted by trained interviewers who used a computer-assisted telephone interview (CATI) technique. As the instrument on use of preventive services included questions on hormone replacement and screening for breast and cervical cancer, only female interviewers were selected. The interviews were monitored by the Italian EUROHIS researchers, who could listen in to any of the interviews at any time. Any problems that arose during the interviews were discussed in debriefing sessions with the interviewers afterwards.

Sample design

301 people could be interviewed within the time and cost (approximately €10 000) constraints. A random two-stage sample was used. The sampling population in the first stage was made up of telephone subscribers, as approximately 92% of the Italian population is registered in the TELECOM file. The first-stage units were stratified by geographical area and demographic size of municipalities, following a uniform allocation

¹ A detailed description of the analysis is available on request.

criterion (the same number of units was allocated in each stratum). This sample design was chosen to take into account the great variability between both the north and the south side of the country and between small and large towns. Within each geographical area, units were selected using random criteria on the basis of quota sampling, defined by linking age groups and sex (50% were male, 50% were female). Only one person in each household was interviewed; he/she answered the questions for him/herself and also for the family. Proxy answers were accepted only for children under 14 years.

Care was taken to achieve a high response rate. The interviews were carried out during suitable hours for families to be contacted (Monday to Friday from 18.30 to 21.30). The distribution of the respondents by gender, age and geographical area is shown in Table 3.3.

Table 3.3. Distribution of pre-test respondents by gender, age and sample stratum, Italy 2000

Strata	Total (N)	Gender		Age group (years)				
		Male	Female	0–14	15–24	25–44	45–64	65+
<i>By geographical area</i>								
North	100	50	50	11	12	31	27	19
Middle	99	49	50	13	11	30	26	19
South	102	50	52	18	16	30	23	15
Total	301	149	152	42	39	91	76	53
<i>By municipality size</i>								
0–250 000 population	151	74	77	22	20	46	42	21
>250 000 population	150	75	75	20	19	45	34	32
Total	301	149	152	42	39	91	76	53

Findings

The questions on chronic physical conditions worked well. The respondents felt that they were reporting on an important aspect of their health. Some problems arose in the case of proxy interviews referring to children, however, where the disease-specific question produced negative feelings in the respondents, resulting in a lack of information about the other items in the check-list or even an interruption of the interview. It was suggested that the order of diseases in the check-list should be further revised for respondents under 15 years old, such as by removing the diseases that could cause anxiety or discomfort on the part of the respondents or their parents (e.g. malignant tumour, heart attack and stroke).

The pre-test revealed a higher prevalence of long-standing health problems in women, and greater prevalence of illness with increasing age. These findings are consistent with the epidemiological literature. The overall prevalence of chronic disease (24.3%, Table 3.4) appeared to be underestimated in the first (open-ended) question, however, as the prevalence estimated from the responses to the disease-specific question was higher, e.g. 29.2% for arthritis/arthrosis alone (Table 3.5). It appeared that some people who considered themselves to be healthy, or who did not consider their condition to be chronic or long-standing, answered “No” at the first question and “Yes” when they recognized their condition in the list.

Table 3.4. Italian pre-test 2000. Open-ended question: Do you have any long-standing illness or health problem?

	Total	Gender		Age group (years)				
		Male	Female	0–14	15–24	25–44	45–64	65+
No. of respondents	301	149	152	42	39	91	76	53
Percentage	100	49.5	50.5	14.0	13.0	30.2	25.2	17.6
Yes	73	31	42	3	2	18	24	26
Percentage of Yes	24.3	20.8	27.6	7.1	5.1	19.8	31.6	49.1

Table 3.5 compares the prevalence of specific diseases in the pre-test conducted in 2000 with the prevalence estimated from the pilot health interview survey conducted in Italy in 1998. Both surveys used the CATI technique and were carried out during the month of October. The disease prevalences in the two studies were generally similar, despite some differences between the two surveys with regard to sample design and size. The prevalence of asthma was difficult to compare because of the different wording used in the two surveys and the different order in which the diseases were listed. In the pilot 1998 HIS survey, allergy was at the top of the list and the question about asthma referred essentially to “bronchial asthma”.

Table 3.5 Data from Italian pre-test 2000 compared with data from pilot Italian HIS, 1998. Prevalence per 100 respondents

	Pre-test 2000	Pilot HIS 1998
1. Asthma	10.3	3.8
<i>If yes, is it allergic asthma?</i>		
2. Allergy (excluding allergic asthma)	14.6	16.7
3. Diabetes	2.7	3.8
4. Cataract	4.3	5.0
5. Hypertension (high blood pressure)	14.6	16.2
6. Heart attack (myocardial infarction)	1.3	3.0
7. Stroke, cerebral haemorrhage	1.3	1.1
8. Chronic bronchitis, emphysema	5.3	7.7
9. Serious skin diseases	6.3	–
10. Arthrosis, (rheumatic) arthritis	29.2	29.0
11. Osteoporosis	8.0	6.6
12. Gastric or duodenal ulcer	5.6	5.8
13. Malignant tumour (including leukaemia and lymphoma)	1.7	2.0
14. Migraine and frequent headache	17.6	14.8
15. Chronic anxiety and depression	13.0	–

The revised order in which the specific diseases were listed (Table 3.6) appeared to be more suitable. No loss of information occurred because of the order of the items, with the exception of proxy interviews referring to children, as mentioned before. The item “Asthma” and the added question “If yes, is it allergic asthma?” revealed that about three-quarters of asthmatic conditions were related to allergy. It would appear that the inclusion or exclusion of allergic asthma affects the estimated prevalence of asthma and allergy. The additional question appeared to solve this problem.

It had previously been suggested that “serious skin diseases” might be a useful disease group to add to the check-list. To investigate this, an open question was added to see what condition people were referring to when they noted that they had “serious skin disease”. Of 19 answers, six could not be related to “serious skin diseases”, which suggested that the term is too general and does not refer to a well defined group of diseases.

Thirty respondents listed a total of 32 diseases in the open-ended question at the bottom of the list. All of these appeared to be chronic in nature, which also lent support to the validity of the question.

Table 3.6. Italian pre-test results: disease-specific question
Do you have or have you ever had any of the following chronic illnesses or conditions?
(reply to each one)

Diseases or conditions	Do you have or have you ever had any of the following conditions?		Was it diagnosed by a physician?		Do you take drugs or are you under therapy?		Did you have this problem in the past 12 months?	
	No	Yes	No	Yes	No	Yes	No	Yes
1. Asthma	270	31	1	30	7	24	17	14
<i>If yes, is it allergic asthma?</i>	8	23						
2. Allergy (excluding allergic asthma)	257	44	7	37	13	31	13	31
3. Diabetes	293	8	0	8	3	5	3	5
4. Cataract	288	13	1	12	8	5	5	8
5. Hypertension (high blood pressure)	257	44	5	39	6	38	12	32
6. Heart attack (myocardial infarction)	297	4	0	4	0	4	3	1
7. Stroke, cerebral haemorrhage	297	4	0	4	0	4	3	1
8. Chronic bronchitis, emphysema	285	16	1	15	5	11	4	12
9. Serious skin diseases	282	19	1	18	5	14	8	11
10. Arthrosis, (rheumatic) arthritis	213	88	14	74	33	55	27	61
11. Osteoporosis	277	24	0	24	10	14	7	17
12. Gastric or duodenal ulcer	284	17	1	16	2	15	9	8
13. Malignant tumour (including leukaemia and lymphoma)	296	5	0	5	1	4	3	2
14. Migraine and frequent headache	248	53	20	33	23	30	6	47
15. Chronic anxiety and depression	262	39	17	22	16	23	19	20
16. Other (specify)	271	30	0	30	3	27	11	19
.....								
.....								
.....								

The question on diagnosis appeared to work well: about 90% of the respondents who had a chronic disease answered that their disease had been diagnosed by a physician. As expected, this percentage was lower for some commonly self-diagnosed diseases such as migraine or anxiety.

The question “Do you take drugs or are you under therapy?” did not work well, with people sometimes asking for clarification about the reference period of the question. It was suggested that a reference period of “the last 12 months” might be appropriate.

The question “Did you have this problem in the past twelve months?” was sometimes understood by the respondent in terms of “the first occurrence of the disease within the past 12 months”. This may produce confusion between measures of prevalence and incidence. To estimate the prevalence within the last 12 months, a better question would be “Have you had this problem in the past 12 months?”

3.7 Revision of the preliminary common instrument

Following the 2000 pre-test, the disease-specific question of the preliminary instrument was modified as follows.

Draft proposal	Revised proposal
Chronic	Long-standing
Did you have this problem in the past 12 months?	Have you had this problem in the past 12 months?
Was it diagnosed by ...?	Was this condition diagnosed by ...?
Do you take drugs or are you under therapy?	For this condition did you take drugs or have you been under therapy in the past 12 months?
Physician	Doctor
Heart attack (myocardial infarction)	Heart attack
Serious skin diseases	Deleted (not a well defined disease group)
Migraine and frequent headache	Migraine or frequent headache
Chronic anxiety and depression	Chronic anxiety or depression

The other changes made were that the question “Have you had this problem in the past 12 months?” was moved so that it came before the question “Did you take drugs or have you been under therapy for this condition in the past 12 months?” Above the columns, the phrase “For this condition” was replaced by the phrase “If Yes, then:”

3.8 EUROHIS field-test 2001

The above revised questions were used in the EUROHIS field-test in 2001/2002. At the time of writing, field-test results were available from Azerbaijan, Croatia, the Czech Republic, France, Germany, Israel, Kyrgyzstan, Latvia, Lithuania, the Republic of Moldova, Romania, the Russian Federation, Slovakia, the former Yugoslav Republic of Macedonia and the United Kingdom.

A review of the field-test country reports confirmed the positive results of the pre-test. In most of the countries, interviewers had no difficulties in the administration of the questions on chronic conditions and gave positive feedback.

3.8.1 Qualitative analysis

There were wording or translation problems in some countries. In particular, the French translation of the disease-specific question was inappropriate, which resulted in the interviewers often changing the wording of the question or answer. It appeared that in both France and the United Kingdom, “chronic” was better understood than “long-standing”.

Some issues concerning the composition of the list of specific conditions arose during the interviewer debriefings. Romania proposed adding “tuberculosis”; the Russian Federation proposed “ischaemic heart disease”, “gastritis”, “diseases of rectum”, “diseases

of kidney” and “diseases of genitourinary system”; and Croatia proposed “thyroid disease”. Romania and the United Kingdom also suggested the elimination of “allergic asthma” from the list, arguing that it should be part of the more general item “asthma” since not every respondent can distinguish the nature of her/his own asthma.

As mentioned earlier, some respondents (in the Czech Republic, Germany, Romania and the United Kingdom) answered negatively to the open-ended question and positively to a condition in the check-list. In these cases, the interviewers/respondents amended the “No” answer that had been given to the open-ended question.

3.8.2 Quantitative analysis

The quantitative analysis on chronic condition indicators was based both on completion rate analysis (the percentage of missing values by gender) and association analysis (the chi-square test statistic in cross-tabulations). In general, the analysis of missing values showed the general willingness of respondents to answer the questions – except for allergic asthma for which the highest missing rate (in France) was 21.4%.

The analysis of associations was limited to valid cells (more than 50% of the cells counted less than 5). No statistics were computed if one of the two variables was a constant or if the reply to one of the two questions was not available. The analysis of valid cells showed that in most of the cases the presence of a specific disease was confirmed by the respondent’s answers in other parts of the interview (most of the valid cells showed a *P* value of less than 0.05).

3.9 Recommendations for a common instrument on chronic physical conditions

The recommended instrument for the collection of data on chronic physical conditions consists of two questions, as presented in Appendix 3.1. This instrument has been developed for use in health or health-related surveys, where data are collected through face-to-face interviews. As it is unlikely that people other than the respondents themselves would be able to answer these questions on chronic conditions, the use of proxy answers is not recommended. While the first open-ended question on long-standing illness would be easy to include in more general social surveys, there is a danger that it will underestimate the prevalence of chronic diseases. It is therefore recommended that the two questions be used in combination. If possible, these questions should not be placed after other questions that require a heavy effort to complete them.

The check-list method for enquiring about specific conditions has been chosen here because it allows a standardized instrument, where each respondent has the same opportunity to report conditions. This gives a more accurate estimation of the prevalence of specific conditions and produces fewer coding problems. The conditions in the list were selected on the basis of their serious nature (they are either potentially fatal and/or limit daily activities) and high economic cost (through the use of health care services, medications, etc.). To avoid underreporting of less severe diseases, the most serious conditions are placed lower down in the check-list.

The order of conditions in the check-list should be the same for all countries and, since all the respondents of a survey should respond to these questions on chronic conditions, there should not be any filter at the start of the question, such as “If yes, then ...which disease”. Countries that wish to collect data on conditions that are not included in the check-list are strongly advised to place them at the bottom of the list, before the open-ended category. It is important always to include this open category at the end in order to

collect data on diseases that have not been named in the list.

It should be noted that the first open-ended question refers to present state of health, while the disease-specific question refers to both present and past health. While only the first answer in the check-list question (Do you have or have you ever had ...? Yes/No) is strictly necessary, the collection of data on the prevalence of conditions within the past 12 months is useful for the purposes of population health monitoring, as it can aid in estimating both the use of and the need for health services. Similarly, the question on diagnosis by a doctor has been included as a more objective indicator of prevalence (people may have incorrectly diagnosed their own disorder), while the question on use of medical or pharmaceutical treatment has been included in order to assess severity of illness.

3.10 Conclusions

The common instrument for chronic conditions proposed here is the result of three years' work. Although many countries include an assessment of chronic physical conditions in their health interview surveys, cross-national comparability of prevalence is very limited owing to the lack of international harmonization of methods. The aim of this work was to develop a common instrument that took into account the experiences that countries already have in assessing chronic illness. Although a detailed review of current practices revealed great variety in the approaches used in different countries, it was possible to identify elements that could be used in the common instrument.

Field-testing of the common instrument revealed positive feedback in terms of the performance of the questions, and led to some changes being made. The final recommended instrument comprises two questions that are typically included in national interview surveys: an open-ended question that inquires whether or not respondents have a chronic condition, and a disease-specific question. The first question allows assessment of overall prevalence of self-perceived chronic conditions, while the second measures the prevalence of specific diseases.

Countries are encouraged to include the recommended instrument on chronic physical conditions in as many national health-related interview surveys as possible. This will provide the opportunity to test the performance of the questions on a larger scale, and will allow the comparability of the resulting data to be investigated.

References

- HELIÖVAARA, M. ET AL. (1993) Reliability and validity of interview data on chronic diseases. *Journal of clinical epidemiology*, **46**: 181–191.
- HUPKENS, C. (1998) *Coverage of health topics by surveys in the European Union*. Brussels, European Commission (Eurostat Working Paper, Population and Social Conditions, 3/1998/E/n.10).
- MAGGI, S. ET AL. (1994) The Italian Longitudinal Study on Aging (ILSA): design and methods. *Aging*, **6**: 464–473.

Chapter 4.

Development of a common instrument for mental health

Howard Meltzer
Office for National Statistics, London, United Kingdom

4.1 Introduction

Work was not started on this indicator until February 2001, owing to staffing changes at the institution that had intended to act as principal investigator for the indicator. The re-established mental health indicator network held its first meeting in London in February 2001. The purpose was to reach a consensus on common instruments for mental health, recognizing that this should be in the context of health interview surveys (HIS) and not surveys of psychiatric morbidity. Initially, the EUROHIS project had aimed at developing a common instrument for “mental disability” but the group decided to depart from this, considering the term somewhat ambiguous. The group decided to replace it with the notion of “mental health” and psychological distress. It was understood that definite recommendations from this first meeting would be needed to fit into the timetable of EUROHIS field-testing later in 2001.

The group members agreed that considerable work had already been done elsewhere on mental health indicators and instruments, and that it was not necessary to start from the beginning to create new instruments. A range of reports, working papers and instruments were used at the meeting as the basis for discussions on the concepts to be included and the development of a draft common instrument.

4.2 Concept exploration

Before considering specific instruments, the group felt it was important to establish the concepts of mental health that should be included in a common instrument. The concepts suggested for discussion were:

- positive mental health
- psychological distress
- role limitation
- social support and social isolation
- assessment of the substantial common psychiatric disorders
- sleep problems
- alcohol and drug dependency
- suicide
- stressful life events
- cognitive functioning and mental health of children.

There was unanimous support for measures of positive mental health and psychological distress.

As the title of the original EUROHIS network related to “mental disability” rather than “mental health”, there was a wide-ranging discussion on all the possible types of disability that could be relevant. Asking questions on activities of daily living (ADL) and instrumental activities of daily living (IADL) seemed attractive, but it was appreciated that this would be too much for a HIS. The group finally concurred that the measurement of social functioning was the most important aspect of disability in this context, incorporating as it did social support and role limitations but excluding social isolation.

Although there was general agreement on the need to measure the prevalence of mental disorders, there was a range of opinion as to how comprehensive this should be. Suggestions included the measurement of one disorder (e.g. depression), two disorders (e.g. anxiety and depression) or several neurotic disorders (e.g. anxiety, depression, panic, obsessive–compulsive disorder, phobias). It was accepted that a HIS is not an appropriate vehicle for obtaining precise estimates of psychiatric disorders, but it was felt that it should include instruments that measure the most common disorders of relevance for public health: anxiety and depression.

Although sleep problems are symptoms of anxiety and depression, the group felt it was a topic that deserved separate consideration.

Although many HIS have questions on alcohol consumption, and some also cover drug taking, the key issue in relation to public health policy is alcohol and drug dependency.

Since suicide is a key public health concern for all European countries, it was considered that a HIS should include questions relating to suicide. The initial suggestion was that there should be one question: whether or not the person had ever attempted suicide. Later, it was felt that frequency of attempts should be included. There was also a suggestion that a question on suicidal thoughts and a measure of their frequency should also be included.

The inclusion of a section on stressful life events was not accepted.

4.3 Selection of measurement instruments

The group then considered the range of instruments that have been used to measure mental health, first on scientific principles (i.e. reliability, validity and acceptability) and second in terms of harmonization with recommendations from other expert working groups in Europe. Finally, the group reviewed the common instrument that could be derived from these existing instruments.

4.3.1 Positive mental health

There is now a general acceptance that surveys that have a section on mental health should include some questions on positive mental health. Although there are four questions in the SF-36 on energy and vitality (Ware & Kosinski 2001), an alternative option for a HIS was the inclusion of the 4-item WHOQOL Positive Feelings measure (WHOQOL Group 1998). The suggestion of adding one question on happiness was proposed and accepted by the group, but later excluded. A further option would have been the WHO-Five wellbeing scale (Bech et al. in press) that is derived from the Psychological General Well-Being Scale. This scale is a unidimensional measure of positive wellbeing, and three of its five items are similar to SF-36 items.

There is no explicit positive mental health scale in the SF-36. The “vitality” scale consists of two positive and two negative items, and does not cover the conceptual range of positive feelings as assessed by the WHOQOL-100 subscale (which consists of four items). However, since there was another EUROHIS group recommending instruments on quality of life, the mental health group resolved to take a pragmatic decision to take the SF-36 items, because these were already in public use. Hence, the four questions of SF-36 were selected.

4.3.2 Psychological distress

The choice was between the Mental Health Indicator 5 (MHI5) derived from the SF-36, the GHQ-12 (Goldberg & Williams 1998; Goodchild & Duncan-Jones 1985) or the K10 (Andrews & Slade 2001).

The majority decision was for the MHI5, as it is shorter and has similar reliability to the GHQ-12. The GHQ-12 has the disadvantage of not covering chronic conditions, has not been validated for the population over 64 years of age, and is copyrighted. Nevertheless, it was noted that the GHQ-12 is the basis for measuring healthy life expectancy in the EURO-REVES network. After the initial meeting, more consideration was given to the K10 questionnaire. However, as this instrument was still under development, and no members of the group had any experience of using it, the group felt that this should not be given priority.

4.3.3 Role limitation

Most HIS have questions on limiting long-standing illness, which is often referred to in the field of psychiatric epidemiology as social impairment. Apart from the European Community Household Panel, most surveys ask two questions, one relating to chronic physical conditions and the other to emotional problems. Therefore, there was a need to add one question on limitation of work or other regular day-to-day activities as a result of emotional problems. It was considered that such questions would fit more appropriately into the WHO recommended instrument on disability (de Bruin et al. 1996) but, if this was not possible, then the three questions from the SF-12 (Ware et al. 1996, 1998; Jenkinson & Layte 1997) on role limitation should be included under mental health.

4.3.4 Social support and social isolation

One of the main disabling consequences of mental health problems is a limitation in social functioning. The term “social functioning” has two distinct aspects: social support and social isolation. The instrument proposed for social support was the 3-item Oslo Scale (Dalgard 1996), which comprises items on primary support group, interest and concern shown by others, and ease of obtaining practical help.

A measure of social isolation was also considered by the group. This instrument comprised four items: someone to confide in, someone to count on in a crisis, someone to count on in personal decision-making, and feeling loved and cared for. In discussions after the meeting, it was felt that these questions could be dropped as they overlapped to some extent with social support.

4.3.5 Common disorders: anxiety and depression

One measure proposed for this domain was the CIDI-Screen (Wittchen et al. 1999), a 12-item questionnaire that takes on average 3½ minutes to complete and gives an estimate of the prevalence of anxiety and/or depression. The other suggestion was to use the screening questions from the revised Clinical Interview Schedule (CIS-R) (Lewis & Pelosi 1990; Lewis et al. 1992). However, the latter is not a recognized screening instrument. There was a proposal that the CIDI-Short Form (CIDI-SF) (Kessler et al. 1998) for Generalized Anxiety Disorder and Major Depressive Episode should be used, which means about 10 questions per disorder. Ultimately, the CIDI-SF was recommended because of its brevity and because it fitted in with the WHO World Mental Health Survey (WHO 1992).

4.3.6 Sleep problems

The questions on sleep from the WHO Health and Health Responsiveness Survey (The WHO World Health Surveys, <http://www3.who.int/whs>) were suggested, i.e. three questions.

4.3.7 Alcohol dependence

The shortest and most widely accepted questionnaire on alcohol dependence is the CAGE (Mayfield et al. 1974). There are many other instruments on alcohol dependency, e.g. AUDIT (Saunders et al. 1993) or questions derived from the US ECA study (Robins & Regier 1991), but the four-item CAGE was generally accepted as preferable for inclusion in HIS.

4.3.8 Suicidal thoughts and behaviour

As specific instruments did not appear to exist for this domain, four new questions were required: (a) whether the respondent had had suicidal thoughts; (b) if yes, the number of times; (c) whether the respondent had made suicidal attempts; and (d) if yes, how many times. Two reference periods were suggested: in the past year or over a lifetime. The final decision was to enquire whether the respondent had ever thought about attempting suicide, or had actually attempted suicide.

4.3.9 Cognitive functioning

Most questions on cognitive functioning are culturally sensitive and cannot be easily applied when aiming for cross-nationally comparative data. A possibility was that, at the beginning of the common instrument, the Mini-Mental Test (Folstein et al. 1975; Nelson et al. 1986; Vertesi et al. 2001) could be asked of all those aged 65 years or over, to decide whether or not to continue the interview. This was a suggestion for further consideration. If there was a desire to include questions on cognitive deficit in HIS, then two items on memory and concentration from the Canadian Health Survey (Swain et al. 1999) were regarded as appropriate.

4.3.10 Children

For children it is important to obtain data on emotional, behavioural and hyperkinetic disorders that have an impact on the child and place a burden on the family. If the HIS is to be designed for children, the short form of Goodman's Strengths and Difficulties Questionnaire (Goodman 1997, 1999; Goodman & Scott 1999; Goodman et al. 1998) should be asked. This has six items: five refer to symptoms, while the last asks about the impact of the symptoms on school and family life.

4.4 Reduction of selected instruments to a minimum set of items

In designing the field-test in which instruments for all eight of the EUROHIS indicators would be tested together, it was necessary to omit some of the mental health items because of the limited time acceptable for telephone interviews. It was noted, however, that this should not be done simply on the basis of the length of the original measure. First, in some of the measures selected, not all items are applicable to all respondents. For example, although the two CIDI instruments for measurement of anxiety and depressive disorders contain around 65 questions, for the 75% of respondents who would be expected to answer negatively to the introductory questions, the total number of questions answered would be about eight. Second, the psychometric properties of many mental health instruments require that all questions have the possibility of being asked. It would then be inappropriate to take only part of the measure.

In selecting which measure(s) to omit, attention was therefore paid to (a) the applicability of a measure (e.g. to all respondents or only to a sub-group), (b) cut-off points or filter questions within the series of questions, (c) the feasibility of omitting some items from the measure, and (d) the ease of answering the questions. Consideration was also given to the possibility of overlap with the other EUROHIS draft instruments, and whether the measures had already been extensively tested elsewhere (and therefore did not require further field-testing).

Within this context, it was decided to omit the following measures:

- positive mental health
- sleep problems
- alcohol dependency
- suicide
- cognitive functioning
- mental health of children.

The following measures were thus retained:

- psychological distress (5 items from the SF-36)
- role limitation (3 items from the SF-36)
- Oslo Social Support Scale (3 items)
- CIDI-Short Form for major depressive episode and generalized anxiety disorder (4 items for respondents with no anxiety or depression within last 12 months; about 18 items for a respondent with depression; about 24 items for a respondent with anxiety).

4.5 Field-testing 2001

The selected mental health instruments were included in the EUROHIS field-test exercise in 2001. The qualitative results (which mostly consist of the feedback from interviewers) relate mainly to the three European Union countries (France, Germany and the United Kingdom) where interviewer feedback was collected systematically and using the same methodology. The instruments that are described in this section are shown in Appendix 4.1.

4.5.1 Psychological distress (from SF-36)

It should be noted that, in the field-test, only five of the six SF-36 response categories were included for these items, i.e. All of the time, Most of the time, Some of the time, A little of the time and None of the time. The middle response category, “A good bit of the time”, was omitted in error.

Qualitative analysis

The only critical remark emerged from the debriefing of the interviewers who carried out the survey in France. They felt that the mixture of positive and negative statements to be rated was sometimes confusing for the respondents. However, such alternation of positive and negative statements is a key part of the construction of an instrument that collects attitudinal data, as it mitigates against “response set”, i.e. it reduces the chance that respondents will give the same answers to all the questions. Essentially, this technique makes the respondent think about each question and the response categories.

The occasions on which the interviewers had difficulties in reading the question or the answer categories were very few – 40 out of the 400 interviews in each country were scrutinized for these types of problems. There were no problems at all in the French or German versions and just one problem was reported on Question 4 (During the past 4 weeks, have you felt downhearted and depressed?) in the English version.

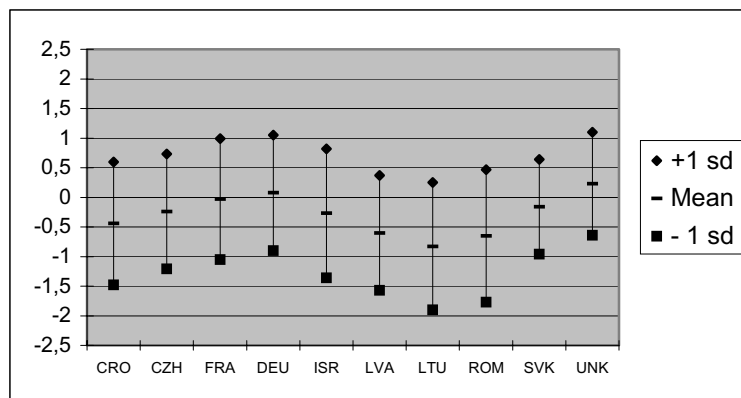
Although one or two of the English interviewers changed either the question wording or the response categories, they did not mention what they felt needed changing and what the change entailed. However, the question wording and response categories are part of an established instrument that has been extensively used and tested, and it is recommended that the wording should not be changed.

Quantitative analysis

Across France, Germany and the United Kingdom hardly any of the 1200 respondents (less than 0.5%) refused to answer the question or said they did not know. The distributions of scores on the five individual items were in the expected range and the apparent gender differences were in the right direction. The results in Fig. 4.1 show the means and standard deviations of the z-transformations of the MHI5 scores across ten field-test countries compared with scores from the US general population (mean of 0 and standard deviation of 1; higher scores represent better mental health).

It should be noted that the EUROHIS field-test surveys were carried out in different ways, despite an attempt at harmonizing the methodology (see Table 4.1 for fuller survey descriptions). Thus, they had different sample sizes and different coverage (areas, municipalities, regions, country), used different samples (e.g. private households, patients visiting the general practitioner), used different sampling strategies and modes of administration (face-to-face, telephone and postal), had different ages of the target population and produced either weighted or unweighted data. The data from these countries cannot therefore be reliably compared and are shown here for illustrative purposes only.

Fig. 4.1. Mean and standard deviation of Z-transformed Mental Health Index score from the SF36 (compared with USA: Mean 0 and Standard Deviation 1)



CRO = Croatia
 CZH = Czech Republic
 FRA = France
 DEU = Germany
 ISR = Israel
 LVA = Latvia
 LTU = Lithuania
 ROM = Romania
 SVK = Slovakia
 UNK = United Kingdom

Conclusion

The MHI5 questions on psychological distress should be part of the core set of items on psychological distress to be included in HIS.

4.5.2 Role limitation (from SF-12)

Qualitative analysis

There were few problems with the question wording or answer categories in English or German, but there were some suggestions for changes in the phrasing of the questions in French. This may have been due to the mode of administration (telephone interviews). These problems should be overcome by clearer instructions to interviewers.

Quantitative analysis

Across France, Germany and the United Kingdom hardly any of the 1200 respondents (less than 1%) refused to answer the questions or said they did not know. The distribution of the answers across the three sampled countries was very similar and of a sensible order of magnitude.

Conclusion

These questions should definitely be included in the core of the mental health module, as they supply the key data on role limitations due to mental health problems. Along with social support questions, they provide essential data on the economic and social consequences of poor mental health.

4.5.3 Social support (Oslo Scale)

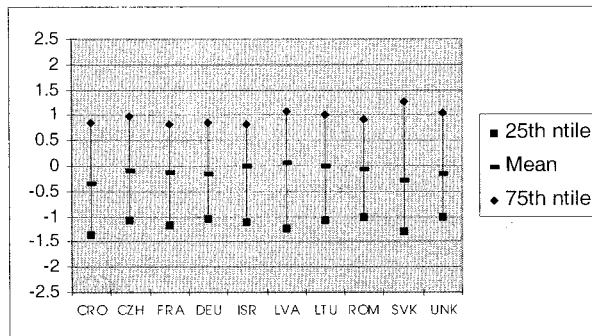
Qualitative analysis

There were no problems with the question wording or answer categories in English or German but there was some concern among the interviewers who asked the questions in French. On several occasions, for all three questions, the French interviewers changed the question wording or answer categories. It was not clear why this was done. The translation has been reviewed since, but the performance of the instrument in France needs further investigation.

Quantitative analysis

Across France, Germany and the United Kingdom hardly any of the 1200 respondents (less than 1%) refused to answer the questions or said they did not know. The distribution of the scores seems reasonable. The data for ten field-testing countries are shown in Fig. 4.2. Again, the different ways in which the surveys have been conducted need to be borne in mind. Fig. 4.2 is therefore included only to illustrate the type of comparison that can be made; the data distribution itself is not particularly relevant here.

Fig. 4.2. Median and 25th and 75th percentiles of a Z-transformation of the Oslo Social Support Score



CRO = Croatia	LVA = Latvia
CZH = Czech Republic	LTU = Lithuania
FRA = France	ROM = Romania
DEU = Germany	SVK = Slovakia
ISR = Israel	UNK = United Kingdom

Conclusion

There should be no problem at all in including this section on social support in the core set of questions in the recommended module on mental health. Because of their general nature, they can also be used as a dependent variable for many other questions in HIS, particularly chronic physical illness and quality of life.

4.5.4 Major depressive episode (CIDI-Short Form)

Qualitative analysis

The general feeling was that this section was repetitive and too long. This probably arose because the measure of psychological distress was asked in the same module as that for major depressive episode, and these measures to some extent cover the same ground. The German interviewers reported that respondents with depression took longer on this section, as they spoke slowly. This is exactly what one would expect as respondents without depression would only be asked a couple of questions, whereas all the questions may apply to a person with depression.

One way to cut down on these questions would be to omit the items that mainly relate to consultations with health professionals and use of medication. These questions do not form part of the diagnostic algorithm for major depressive episode and are probably more appropriate for surveys of psychiatric morbidity than HIS, which cover a whole range of issues.

The occasions on which the interviewers had difficulties in reading the question or the answer categories in English, French or German were very few. Some interviewers questioned the words “sad” and “blue” in the first item as being rather too American. However, the three words in the item, “sad, blue or depressed”, are shown as alternatives and there should be no difficulty in understanding the concept.

Quantitative analysis

Across all three countries, hardly any of the 1200 respondents (less than 1%) refused to answer the question or said they did not know. As expected, the data also demonstrated that more than 80% of the three samples answered a maximum of three questions. The prevalence of major depressive episode in the past year was found to be about 12% in the United Kingdom and 6% in both France and Germany. The rate for the United Kingdom is far higher than expected. Data from other countries that participated in the field test are shown in Table 4.1.

The rate for Lithuania was particularly high, and this may be because it was the only field-test that used a postal survey. However, the above-mentioned limitations due to different methods of administration apply.

Table 4.1. Twelve-month prevalence rates of major depression

Country	12-month prevalence	Sample size	Age range (years)	Weighted or unweighted data	Type of sample	Mode of administration
Croatia	10.0	361	16–85	Unweighted	GP register	Face-to-face
Czech Republic	7.6	512	15+	Unweighted	Population register	Face-to-face
France	6.3	400	35+	Weighted	RDD ^a	Telephone
Germany	5.5	402	35+	Weighted	RDD ^a	Telephone
Israel	7.9	992	16+	Unweighted	RDD ^a	Telephone
Latvia	7.4	323	16+	Unweighted	Population register	Face-to-face
Lithuania	17.1	455	20–64	Unweighted	Population register	Mail survey
Romania	11.8	602	20–79	Unweighted	Population register	Face-to-face
Slovakia	11.0	306	16–75	Unweighted	Convenience sample from region	Face-to-face and self-administered
United Kingdom	11.9	402	35+	Weighted	RDD ^a	Telephone

^a RDD = A random sample of telephone numbers of the national telephone company.

Conclusion

The items for major depression (without those looking at service use) of the CIDI-Short Form should be recommended as an optional set of questions in HIS when there is a particular interest in the prevalence of mental disorders.

4.5.5 Generalized anxiety disorder (CIDI-Short Form)

Qualitative analysis

There were very few problems with this section in terms of question wording. One or two changes were made by interviewers, e.g. the word “period” in the second item should be “time period” to avoid any ambiguity. However, the general comment that the whole section on mental health was too long and too repetitive needs to be taken into account. As with the questions on major depressive episode, one way of cutting down on these questions would be to omit the items that mainly relate to consultations with health professionals and use of medication. As in the depression measure, these questions do not form part of the diagnostic algorithm for generalized anxiety disorder and are probably more appropriate for surveys of psychiatric morbidity than HIS.

Quantitative analysis

Across France, Germany and the United Kingdom, hardly any of the 1200 respondents (less than 1%) refused to answer the question or said they did not know. As expected, the data also demonstrated that more than 70% of the three samples answered a maximum of two questions.

The prevalence of generalized anxiety disorder in the past year was around 5% in France and the United Kingdom and about 3% in Germany. These values are plausible. Data from ten countries that participated in the field-test are shown in Table 4.2. Again, these data should be compared with caution, bearing in mind the different methodological approaches used.

Table 4.2 Twelve-month prevalence rates of generalized anxiety disorder

Country	12-month prevalence	Sample size	Age range (years)	Weighted or unweighted data	Type of sample	Mode of administration
Croatia	7.2	361	16–85	Unweighted	GP register	Face-to-face
Czech Republic	4.7	512	15+	Unweighted	Population register	Face-to-face
France	4.6	400	35+	Weighted	RDD ^a	Telephone
Germany	2.6	402	35+	Weighted	RDD ^a	Telephone
Israel	2.3	992	16+	Unweighted	RDD ^a	Telephone
Latvia	3.1	323	16+	Unweighted	Population register	Face-to-face
Lithuania	16.0	455	20–64	Unweighted	Population register	Mail survey
Romania	4.3	602	20–79	Unweighted	Population register	Face-to-face
Slovakia	5.3	306	16–75	Unweighted	Convenience sample from region	Face-to-face and self-administered
United Kingdom	5.3	402	35+	Weighted	RDD ^a	Telephone

^a RDD = A random sample of telephone numbers of the national telephone company.

Conclusion

The items for generalized anxiety disorder (without those looking at onset or service use) should be recommended as an optional set of questions in HIS when there is a particular interest in the prevalence of mental disorders. Again, the result for Lithuania seemed very high and may be due to the survey being carried out by post. One would expect the results from Croatia to be the highest, as the sample here consisted of patients visiting the general practitioner.

4.6 Recommended common instruments for mental health

The recommended common survey instruments have been divided into three sets. The first represents measures that are recommended as core indicators on mental health, i.e. those that are regarded as key to understanding the mental health of the population in general terms. For this core set, four instruments are proposed although only three were field-tested: positive mental health, psychological distress, role limitation and perceived social support. It is recommended that the four instruments be asked in the order shown in the description of the instruments in Appendix 4.1. There are three main advantages of asking these four instruments together.

- Three of the instruments come from the SF-36, which is already used in many European countries, and there are normative data with which to make comparisons. For those countries that already use the SF-12, the addition of some extra questions is less onerous than having to add a whole new set of questions.
- All instruments take a dimensional rather than a categorical approach to the measurement of the underlying concept, thus producing a distribution of scores across the population.
- The total number of questions for all four instruments is 15 and should take no more than 5 minutes.

The second set of recommended survey instruments comprises two instruments that also

come from the same much larger instrument: the CIDI-SF. They measure major depressive episode and general anxiety disorder. These two instruments are compatible with the WHO Mental Health Survey (<http://www.who.int/evidence/whs>). They should be used when the focus is on prevalence of mental disorders and where the country does not have a programme of surveys on psychiatric morbidity. Although they appear lengthy for a HIS that covers a myriad of topics, it should be stressed that 70–80% of all survey respondents would only be asked three questions on depression and two on anxiety.

The third set of instruments is optional (i.e. they are desirable but not essential). They cover sleep problems, alcohol dependence and attempted suicide. It is recommended that the alcohol dependence questions be included in the EUROHIS module on use of alcohol.

All these sets of instruments are described in Appendix 4.1, together with guidance on scoring.

References

- ANDREWS, G. & SLADE, T. (2001) Interpreting scores on the Kessler Psychological Distress Scale (K10). *Australia and New Zealand journal of public health*, **25**: 494–497.
- BECH, P. ET AL. (in press) Measuring well-being rather than absence of distress symptoms. A comparison of the SF-36 Mental Health subscale and the WHO-Five well-being scale. *International journal of methods in psychiatric research*.
- DALGARD, O.S. (1996) Community health profile as tool for psychiatric prevention. In: Trent, D.R. & Reed, C., ed. *Promotion of mental health*, Vol. 5. Aldershot, Avebury.
- DE BRUIN, A. ET AL. ED. (1996) *Health interview surveys: towards international harmonization of methods and instruments*. Copenhagen, WHO Regional Office for Europe (WHO Regional Publications, European Series, No. 58).
- FOLSTEIN, M.F. ET AL. (1975) Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, **12**: 189–198.
- GOLDBERG, D.P. & WILLIAMS, P. (1998) *Users' guide to the General Health Questionnaire*. Windsor, NFER-Nelson.
- GOODCHILD, M.E. & DUNCAN-JONES, P. (1985) Chronicity and the General Health Questionnaire. *British journal of psychiatry*, **146**: 55–61.
- GOODMAN, R. (1997) The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry*, **38**: 581–586.
- GOODMAN, R. (1999) The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of child psychology and psychiatry*, **40**: 791–801.
- GOODMAN, R. & SCOTT, S. (1999) Comparing the Strengths and Difficulties Questionnaire and the Child Behaviour Checklist: is small beautiful? *Journal of abnormal child psychology*, **27**: 17–24.
- GOODMAN, R. ET AL. (1998) The Strengths and Difficulties Questionnaire: a pilot study on the validity of the self-report version. *European child and adolescent psychiatry*, **7**: 125–130.
- JENKINSON, C. & LAYTE, R. (1997) Development and testing of the UK SF-12. *Journal of health services research policy*, **2**: 14–18.
- KESSLER, R.C. ET AL. (1998) The World Health Organization's Composite International Diagnostic Interview Short-Form (CIDI-SF). *International journal of methods in psychiatric research*, **7**: 172–185.
- LEWIS, G. & PELOSI, A.J. (1990) *Manual of the revised clinical interview schedule (CIS-R)*. London, Institute of Psychiatry.
- LEWIS, G. ET AL. (1992) Measuring psychiatric disorder in the community: a standardised instrument for use by lay interviewers. *Psychological medicine*, **22**: 465–486.
- MAYFIELD, D. ET AL. (1974) The CAGE questionnaire: validation of a new alcoholism screening instrument. *American journal of psychiatry*, **131**: 1121–1123.
- NELSON, H.E. ET AL. (1986) Bedside cognitive screening instruments: a critical assessment. *Journal of nervous and mental disease*, **174**: 73–83.
- ROBINS, L.N. & REGIER, D.A., ED. (1991) *Psychiatric disorders in America: the Epidemiologic Catchment Area Study*. New York, Free Press.

- SAUNDERS, J.B. ET AL. (1993) Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption – II. *Addiction*, **88**: 791–804.
- SWAIN, I. ET AL. (1999) The national population survey. *Health reports, Statistics Canada*, **10**: 69–82.
- VERTESI, A. ET AL. (2001) Standardized Mini-Mental State Examination: use and interpretation. *Canadian family physician*, **47**: 2018–2023.
- WARE, J.E. & KOSINSKI, M.A. (2001) *SF-36: Physical and Mental Health Summary Scales. A manual for users of Version 1 (2nd edition)*. Lincoln, RI, Quality Metric Inc.
- WARE, J.E. ET AL. (1996) A 12-item Short Form Health Survey (SF-12): construction of scales and preliminary tests of reliability and validity. *Medical care*, **32**: 220–233.
- WARE, J.E. ET AL. (1998) *SF-12: How to score the SF-12 Physical and Mental Health Summary Scales*, 3rd ed. Lincoln, RI, Quality Metric Inc. and Boston, MA, Health Assessment Laboratory.
- WHOQOL GROUP (1998) The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. *Social science and medicine*, **46**: 1569–1585.
- WITTCHEN, H.U. ET AL. (1999) Screening for Mental Disorders: performance of the Composite International Diagnostic Screener (CID-S). *International journal of methods in psychiatric research*, **8**: 59–70.
- WORLD HEALTH ORGANIZATION (1992) *The ICD-10 Classification of Mental and Behavioural Disorders: clinical descriptions and diagnostic guidelines*. Geneva, World Health Organization.

Appendix 4.1

EUROHIS recommended common instrument for mental health

1. Core set of recommended instruments: general measures (Except for the SF-36 scale for Positive mental health, these measures were field-tested in 2001 as part of the EUROHIS project)

1.1 Positive mental health (from SF-36)

How much, during the past 4 weeks....

		All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time
vital1	Did you feel full of life?	1	2	3	4	5	6
vital2	Did you have lots of energy?	1	2	3	4	5	6
vital3	Did you feel worn out?	1	2	3	4	5	6
vital4	Did you feel tired?	1	2	3	4	5	6

Scoring Psychological wellbeing (Positive mental health or Vitality) from the SF-36

* THE SF-36 VITALITY ITEMS (vital1 to vital4).

* REVERSE TWO ITEMS (vital1 and vital2).

* AFTER ITEM REVERSAL, ALL ITEMS ARE POSITIVELY SCORED i.e. the higher the score, the less the fatigue and the greater the energy

* THIS SCALE IS POSITIVELY SCORED i.e. the higher the score, the greater the vitality

SPSS programme for calculating level of psychological wellbeing

i.e. Vitality index: vital

DO REPEAT i = vital1,vital2,vital3,vital4.

+ IF (i < 1 OR i > 6) i = \$SYSMIS.

END REPEAT.

COMPUTE vital = 100*(MEAN.2(7-vital1,7-vital2,vital3,vital4) - 1)/5.

VARIABLE LABEL vital 'SF-36 Vitality index (0-100)'.

1.2 Psychological distress (from SF-36)

How much, during the past 4 weeks....

	All of the time	Most of the time	A good bit of time	Some of the time	A little of the time	None of the time
mhi1 Did you feel very nervous?	1	2	3	4	5	6
mhi2 Have you felt so down in the dumps, nothing could cheer you up?	1	2	3	4	5	6
mhi3 Have you felt calm and peaceful?	1	2	3	4	5	6
mhi4 Have you felt down-hearted and depressed?	1	2	3	4	5	6
mhi5 Have you been very happy?	1	2	3	4	5	6

Scoring Psychological distress (Mental Health Index, MHI) from the SF-36

* THE SF-36 MENTAL HEALTH ITEMS (mhi1 to mhi5)

* REVERSE TWO ITEMS (mhi3 and mhi5)

* AFTER ITEM REVERSAL, ALL ITEMS ARE POSITIVELY SCORED i.e. the higher the score, the better the mental health

* THIS SCALE IS POSITIVELY SCORED i.e. the higher the score, the better the mental health

SPSS programme for calculating level of psychological distress.

i.e. Mental Health Index (mhi)

DO REPEAT i = mhi1,mhi2,mhi3,mhi4,mhi5.

+ IF (i < 1 OR i > 6) i = \$\$SYSMIS.

END REPEAT.

COMPUTE mhi = 100*(MEAN.3(mhi1,mhi2,7-mhi3,mhi4,7-mhi5) - 1)/5.

VARIABLE LABEL mhi 'SF-36 Mental health index (0-100)'.

1.3 Role Limitation (from SF-36)

During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

	Yes	No
rolem1 Cut down on the amount of time you spent on work or other activities?	1	2
rolem2 Accomplished less than you would like?	1	2
rolem3 Didn't do work or other activities as carefully as usual?	1	2

1.4 Perceived social support (Oslo scale)

	Score
Oslo1 How many people are so close to you that you can count on them if you have serious personal problems?	
None	1
1 or 2	2
3 to 5	3
6 or more	4
Oslo2 How much concern do people show in what you are doing?	5
A lot of concern and interest	4
Some concern and interest	3
Uncertain	2
Little concern and interest	1
No concern and interest	
Oslo3 How easy is it to get practical help from neighbours if you should need it?	
Very easy	5
Easy	4
Possible	3
Difficult	2
Very difficult	1

Description of scale

A total score is calculated by summarising the standardised Z scores for each item.

The total score can be used as a continuous variable, or as a categorical variable, split for instance into 3 or 5 groups of equal size.

SPSS programme for calculating Overall score

```
IF (oslo1 <1 or oslo1 > 4) oslo1 =$$SYSMIS.
```

```
IF (oslo2 <1 or oslo2 > 5) oslo2 =$$SYSMIS.
```

```
IF (oslo3 <1 or oslo3 > 5) oslo3 =$$SYSMIS.
```

```
DESCRIPTIVES
```

```
  VARIABLES=oslo1 oslo2 oslo3 /SAVE/MISSING=LISTWISE .
```

```
  /STATISTICS=MEAN STDDEV MIN MAX.
```

```
COMPUTE OSLO = zoslo1 + zoslo2 + zoslo3.
```

2. Set of recommended instruments on prevalence of mental disorders (Both these CIDI instruments were field-tested in 2001 as part of the EUROHIS project)

2.1 Major depressive episode (CIDI-Short Form)

A1. During the past 12 months, was there ever a time when you felt sad, blue, or depressed for two weeks or more in a row?

Yes	1	→ A1a
No	5	→ A9
Volunteered: I was on medication/antidepressants	6	→ A9

A1a. For the next few questions, please think of the two-week period during the past 12 months when these feelings were worst. During that time did the feelings of being sad, blue, or depressed usually last...

all day long,	1	→ A1b
most of the day,	2	→ A1b
about half the day,	3	→ A1b
or less than half the day?	4	→ A9

A1b. During those two weeks, did you feel this way...

every day,	1	→ A1c
almost every day,	2	→ A1c
or less often?	3	→ A9

A1c. During those two weeks did you lose interest in most things like hobbies, work, or activities that usually give you pleasure?

Yes	1
No	5

A1d. Thinking about those same two weeks, did you feel more tired out or low on energy than is usual for you?

Yes	1
No	5

A2. Did you gain or lose weight without trying, or did you stay about the same? (*Interviewer: If Respondent asks: "Are we still talking about the same two weeks?" Answer: "Yes."*)

Gained weight	1	→ A2a
Lost weight	2	→ A2a
If volunteered: lost and gained weight	3	→ A2a
Stayed about the same	4	→ A3
If volunteered: was on a diet	5	→ A3

A2a. About how much did (you gain/you lose/your weight change)? (*Interviewer: Accept a range response*)

----- pounds OR ----- Kilos

A2b. (*Interviewer code: Did R's weight change by 10 pounds / 5 kilos or more?*)

Yes	1
No	5

A3. Did you have more trouble falling asleep than you usually do during those two weeks?

Yes	1	→ A3a
No	5	→ A4

A3a. Did that happen...

every night,	1
nearly every night,	2
or less often during those two weeks?	3

A4. During those two weeks, did you have a lot more trouble concentrating than usual? (Interviewer: If R asks: "Are we still talking about the same two weeks?" Answer: "Yes.")

Yes	1
No	5

A5. People sometimes feel down on themselves, no good, or worthless. During that two-week period, did you feel this way? (Interviewer: If R asks: "Are we still talking about the same two weeks?" Answer: "Yes.")

Yes	1
No	5

A6. Did you think a lot about death -- either your own, someone else's, or death in general during those two weeks?

Yes	1	→ A9
No	5	→ A9

[Items A7 and A8 are not used here]

A9. During the past 12 months, was there ever a time lasting two weeks or more when you lost interest in most things like hobbies, work, or activities that usually give you pleasure?

Yes	1	→ A9a
No	5	→ End
If volunteered: I was on medication/antidepressants	6	→ End

A9a. For the next few questions, please think of the two-week period during the past 12 months when you had the most complete loss of interest in things. During that two-week period, did the loss of interest usually last...

all day long,	1	→ A9b
most of the day,	2	→ A9b
about half the day,	3	→ A9b
or less than half the day?	4	→ End

A9b. Did you feel this way every day, almost every day or less often during those two weeks?

every day,	1	→ A9c
almost every day,	2	→ A9c
or less often?	3	→ End

A9c. During those two weeks did you feel tired out or low on energy than is more usual for you?

Yes	1
No	5

A10. Did you gain or lose weight without trying, or did you stay about the same?
(Interviewer: *If R asks: "Are we still talking about the same two weeks?" Answer: "Yes."*)

Gained weight	1	→ A10a
Lost weight	2	→ A10a
If volunteered: lost and gained weight	3	→ A10a
Stayed about the same	4	→ A11
If volunteered: was on a diet	5	→ A11

A10a. About how much did (you gain/you lose/your weight change)? (Interviewer: *Accept a range response*)

----- pounds OR ----- Kilos

A10b. (Interviewer code: *Did R's weight change by 10 pounds/5 kilos or more?*)

Yes	1
No	5

A11. Did you have more trouble falling asleep than you usually do during those two weeks?

Yes	1	→ A11a
No	5	→ A12

A11a. Did that happen...

every night,	1
nearly every night,	2
or less often during those two weeks?	3

A12. During those two weeks, did you have a lot more trouble concentrating than usual? (Interviewer: If R asks: "Are we still talking about the same two weeks?" Answer: "Yes.")

Yes	1
No	5

A13. People sometimes feel down on themselves, no good, or worthless. During that two-week period, did you feel this way? (Interviewer: If R asks: "Are we still talking about the same two weeks?" Answer: "Yes.")

Yes	1
No	5

A14. Did you think a lot about death -- either your own, someone else's, or death in general during those two weeks?

Yes	1
No	5

SPSS programme for calculating caseness of Major Depressive episode (MD).

*** Creating variable for probable dysphoric depression.**

COMPUTE MD1=0.

* Dealing with missing values.

```
DO IF ((A1 < 1) OR (A1=2) OR (A1=3) OR (A1=4) OR (A1>6))
  OR (A1a < 1) OR (A1a > 4)
  OR (A1b <1) OR (A1b >3).
COMPUTE MD1= $SYSMIS.
END IF.
```

* Counting number of symptoms.

COUNT MD1SYMP = A1c A1d A2b A3a A4 A5 A6 (1).

* Creating first depression variable.

```
DO IF (A1 = 1)
  AND ((A1a = 1) OR (A1a = 2) OR (A1a = 3))
  AND ((A1b = 1) OR (A1b = 2))
  AND (MD1SYMP >2).
COMPUTE MD1 = 1.
END IF.
```

*** Creating variable for probable anhedonic depression.**

COMPUTE MD2=0.

* Dealing with missing values.

```
DO IF ((A9 < 1) OR (A9=2) OR (A9=3) OR (A9=4) OR (A9>6))
  OR ((A9a < 1) OR (A9a > 4))
  OR ((A9b < 1) OR (A9b > 3)).
COMPUTE MD2= $SYSMIS.
END IF.
```

* Counting number of symptoms.

```
COUNT MD2SYMP = A9c A10b A11a A12 A13 A14 (1).
```

* Creating second depression variable.

```
DO IF (A9 = 1)
  AND ((A9a = 1) OR (A9a = 2) OR (A9a = 3))
  AND ((A9b = 1) OR (A9b = 2))
  AND (MD2SYMP > 2).
COMPUTE MD2 = 1.
END IF.
```

* Creating variable for any probable major depressive episode

```
COMPUTE MD=0.
DO IF (MD1 =1 OR MD2 =1)
COMPUTE MD=1.
END IF.
```

VARIABLE LABEL MD "Probable caseness for major depressive episode".
VALUE LABEL MD (1) Probable Caseness (0) Non-probable caseness.

2.2 Generalised Anxiety Disorder (CIDI-Short Form)

B1. During the past 12 months, did you ever have a period lasting one month or longer when most of the time you felt worried, tense, or anxious?

Yes	1	→ B2
No	5	→ B1a

B1a. People differ a lot in how much they worry about things. Did you have a time in the past 12 months when you worried a lot more than most people would in your situation?

Yes	1	→ B2
No	2	→ End

B2. Has that period ended or is it still going on?

Ended	1	→ B2a
Still going on	2	→ B2b

B2a. How many months or years did it go on before it ended?

----- months, or	→ B3
----- years, or	→ B3
If volunteered "all my life" or "as long as I can remember"	→ B3

B2b. How many months or years has it been going on?

----- months, or	→ B3
----- years, or	→ B3
If volunteered "all my life" or "as long as I can remember"	→ B3

B3. *Interviewer check:*

<i>B2a/B2b is 6 months or longer, or volunteered "all my life" or "as long as I can remember"</i>	1 → B4
<i>B2a/B2b is less than 6 months</i>	2 → End

B4. (During that period, was your/is your) worry stronger than in other people?

Yes	1
No	5

B5. (Did/Do) you worry most days?

Yes	1
No	5

B6. (Did/Do) you usually worry about one particular thing, such as your job security or the failing health of a loved one, or more than one thing?

One thing	1
More than one thing	2

B7. (Did/Do) you find it difficult to stop worrying?

Yes	1
No	5

B8. (Did/Do) you ever have different worries on your mind at the same time?

Yes	1
No	5

B9. How often (was/is) your worry so strong that you (couldn't/can't) put it out of your mind no matter how hard you (tried/try) –

often,	1
sometimes,	2
rarely,	3
or never?	4

B10. How often (did/do) you find it difficult to control your worry –

often,	1
sometimes,	2
rarely,	3
or never?	4

B11. What sort of things (did/do) you mainly worry about? (*PROBE: Any other main worries?*)

B12. When you (are/were) worried or anxious,

a. (Are/Were) you restless?

Yes	1
No	5

b. (Are/Were) you keyed up or on edge?

Yes	1
No	5

c. (Are/Were) you easily tired?

Yes	1
No	5

d. (Do/Did) you have difficulty keeping your mind on what you were doing?

Yes	1
No	5

e. (Are/Were) you more irritable than usual?

Yes	1
No	5

f. (Do/Did) you have tense, sore or aching muscles

Yes	1
No	5

g. (Do/Did) you have trouble falling asleep or staying asleep?

Yes	1
No	5

SPSS programme for calculating caseness of General Anxiety Disorder (GAD) (non-hierarchical)

COUNT GADSYMP = B12a B12b B12c B12d B12e B12f B12g (1).

DO IF ((B1 < 1) OR (B1=2) OR (B1=3) OR (B1=4) OR (B1 > 5))

AND ((B1a < 1) OR (B1a > 2))

AND ((B3 < 1) OR (B3 > 2))

COMPUTE GAD = \$SYSMIS.

ELSE IF ((B1 = 1) OR (B1a = 1) AND (B3 = 1))

AND (B4 = 1)

AND (B5 = 1)

AND ((B6 = 2) OR (B8 = 1))

AND ((B7 = 1) OR (B9 = 1) OR (B10 = 1))

AND (GADSYMP > 2).

COMPUTE GAD = 1.

ELSE.

COMPUTE GAD=0.

END IF.

VARIABLE LABEL GAD "Caseness for non-hierarchical generalised anxiety disorder".

VALUE LABEL MD (1) Caseness (0) Not a case.

3. Optional set of recommended instruments

(None of these measures were field-tested in 2001 as part of the EUROHIS project)

3.1 Sleep problems (from WHO Health and Responsiveness survey)

How much time, during the past 4 weeks, did you have a problem with ...

	None of the time	Some of the time	A good bit of the time	Most of the time	All of the time
spi1 Falling asleep?	5	4	3	2	1
spi2 Waking up frequently during the night?	5	4	3	2	1
spi3 Waking up too early in the morning?	5	4	3	2	1

Comment on scoring this measure

The responses to these items have not been made into an index. However, they have been used as a predictor of insomnia (sleep disorders) in primary care samples. To make this measure compatible with the SF-36 measures, a score can be calculated as follows, where '0' signifies severe sleep problems and '100' no sleeping problems.

(Provisional) scoring of Sleep problems

* THE SLEEP ITEMS (spi1 to spi3)

* ALL ITEMS ARE POSITIVELY SCORED – the higher the score, the better the sleep pattern

* THIS SCALE IS POSITIVELY SCORED – the higher the score, the better the sleep pattern

SPSS programme for calculating level of Sleep Problems Index (spi)

```
-----
DO REPEAT i = spi1,spi2,spi3.
+ IF (i < 1 OR i > 5) i = $$SYSMIS.
END REPEAT.

COMPUTE spi = 100*( MEAN.2(spi1,spi2,6spi3) - 1 )/4.

VARIABLE LABEL spi 'Sleep problem index (0-100)'.
```

3.2 Alcohol dependence (CAGE questionnaire)

	Yes	No
F1 Have you tried to cut down your drinking?	1	0
F2 Have the people close to you been annoyed by your drinking?	1	0
F3 Have you had feelings of guilt due to drinking?	1	0
F4 Have you needed an eye-opener in the morning?	1	0

Scoring for CAGE

Sum the responses of F1 to F4.
Minimum score is 0, maximum score is 4

SPSS programme for calculating CAGE score

```
DO REPEAT i = F1,F2,F3,F4.
+ IF (i < 0 OR i > 1) i = $$SYSMIS.
END REPEAT.
```

```
COMPUTE CAGE = (F1 + F2 + F3 + F4).
```

Interpretation of scores

A score of 1 or 2 is indicative of a problem drinker.
A score of 3 or 4 is indicative of alcohol dependence.

3.3 Suicidal thoughts and behaviour

G1	Have you ever thought of killing yourself?		
	Yes	1	→ G2
	No	2	→ End
G2	How many times have you thought of killing yourself?		
	Once	1	
	Twice	2	
	Three times	3	
	Four or more times	4	
G3	Have you ever tried to kill yourself?		
	Yes	1	→ G4
	No	2	→ End

G4 How many times have you tried to kill yourself?

Once	1
Twice	2
Three times	3
Four or more times	4

Scoring the questions on suicidal ideation and behaviour

Two measures are derived from these four items: i) the lifetime prevalence of suicidal ideation, and ii) the lifetime prevalence of suicidal attempts.

SPSS programme for calculating index of suicidal ideation (suicidea)

```
DO IF (G1 < 1 OR > 2).
COMPUTE suicidea= $SYSMIS.
ELSE IF (G1=2).
COMPUTE suicidea=0.
ELSE IF (G2=1).
COMPUTE suicidea=1.
ELSE IF (G2=2).
COMPUTE suicidea=2.
ELSE IF (G2=3).
COMPUTE suicidea=3.
ELSE IF (G2=4).
COMPUTE suicidea=4.
ELSE.
END IF.
```

VARIABLE LABEL suicidea 'Index of lifetime suicidal ideation'.
VALUE LABEL suicidea (0) never (1) once (2) twice (3) 3 times (4) 4+ times.

SPSS programme for calculating index of suicidal attempts (suicatt)

```
DO IF (G3 < 1 OR > 2).
COMPUTE suicatt= $SYSMIS.
ELSE IF (G3=2).
COMPUTE suicatt=0.
ELSE IF (G4=1).
COMPUTE suicatt=1.
ELSE IF (G4=2).
COMPUTE suicatt=2.
ELSE IF (G4=3).
COMPUTE suicatt=3.
ELSE IF (G4=4).
COMPUTE suicatt=4.
ELSE.
END IF.
```

VARIABLE LABEL suicatt 'Index of lifetime suicide attempts'.
VALUE LABEL suicatt (0) never (1) once (2) twice (3) 3 times (4) 4+ times.

Chapter 5.

Development of a common instrument for alcohol consumption

Jussi Simpura

National Research and Development Centre for Welfare and Health (STAKES), Finland¹

5.1 Introduction

Health interview surveys (HIS) nearly invariably include measurement of alcohol consumption, thus reflecting the importance of this topic for population health. The aim is usually to determine the proportion of people with elevated alcohol-related health risk in different population groups and periods of time, and to follow the changes in these proportions. Three aspects of alcohol consumption are most relevant:

- the volume of alcohol consumption;
- the frequency of occasions of high alcohol intake (“binge drinking”); and
- drinking in risky contexts (road traffic, work, certain types of leisure activity).

These issues relate to the two main currents of preventive work concerning alcohol problems: control of consumption (Edwards et al. 1994) and harm reduction (Rehm 1999). To support these goals, the WHO Global Burden of Disease project has included a specific section on alcohol-related harm and its determinants (Rehm & Gmel 2000). The European project on drinking patterns among women (Knibbe & Bloomfield 2001) offers another example whereby existing data sets on drinking patterns are used to conduct comparative analyses across countries.

It is apparent from the “survey of surveys” that was conducted as part of the EUROHIS project, and from a compilation of research on drinking patterns (Simpura & Karlsson 2001a) that was produced in connection with the European Comparative Alcohol Study (ECAS) (Norström 2002), that HIS seldom cover satisfactorily any of the three aspects of alcohol consumption that are mentioned above. Indeed, alcohol consumption typically receives only minor attention, with survey questions tending to focus more on trivial aspects of drinking such as abstinence, overall drinking frequency and perhaps frequency of drinking specific types of beverage. Of course, these aspects also have epidemiological importance (it is known, for instance, that abstainers may have lower or higher risks of cardiovascular diseases than light drinkers), but this is small compared to that of the three main aspects of consumption: volume, binge drinking and risky contexts. Overall drinking frequency can be used as a proxy for the volume of consumption within one country, but not in international comparisons. Beverage preferences (e.g. wine vs. beer vs. distilled beverages) have a much smaller role in explaining the variation of alcohol-related harm than commonly believed (Norström 2001).

¹ Now at Statistics Finland, Helsinki, Finland.

5.1.1 Methodology

The development of the EUROHIS instrument on alcohol consumption proceeded slightly differently from the original plan presented in the EUROHIS protocol, as the results of the survey of surveys became available too late for the preparation of the preliminary instrument. This was not a major problem, however, as the research team had access to other compilations of questionnaires on drinking patterns.

In developing the EUROHIS instrument for alcohol consumption, the emphasis was not on its psychometric properties but rather on the major obstacles that arise in comparative surveys on alcohol use. The following steps were undertaken:

1. development of a draft common instrument in 1999;
2. the “survey of surveys”, complemented with materials from ECAS;
3. protocol analysis: a qualitative investigation of cultural differences in interpreting survey questions;
4. implementation of the draft instrument as part of ECAS in 2000; and
5. field-testing in 2001 alongside the other EUROHIS draft instruments.

5.2 Development of a draft common instrument

A draft common instrument was prepared in 1999 on the basis of the alcohol group’s long experience in international comparative research on drinking patterns.² It was decided that four main aspects should be covered by the instrument: abstinence and prevalence of use, frequencies of drinking and typical volumes, binge drinking, and drinking in specific contexts. This draft instrument was translated into six languages (with no back translations).

An important additional input for the EUROHIS alcohol instrument was provided by the scientific workshop on measurement issues in alcohol consumption and alcohol-related problems, organized by the University of Stockholm in April 2000 (Dawson & Room 2000; Room 2000). An unofficial detailed recommendation³ was prepared for a minimum data set in surveys where time is extremely limited. The following aspects were recommended for inclusion:

- abstinence: lifetime and past 12 months;
- overall frequency of drinking (any beverages; in days per year, month or week);
- usual quantity (any beverages; quantity per day); and
- drinking five drinks or more per day, i.e. the equivalent of 60 grams of ethanol (in days per year, month or week).

² The problems of international comparisons of drinking habit surveys have been regularly discussed at the Kettil Bruun Society’s annual Alcohol Epidemiology Symposia since the 1980s. At these events, informal compilations and reviews of survey questionnaires have also been made. Some members of the EUROHIS alcohol consumption network have actively participated in this process since its inception.

³ Another recommendation is in preparation in *Gender, Alcohol and Culture: an International Study (GENACIS)*. See <http://www.icap.org/international/gender.html> for further information.

Probably the most disputable element here is the use of days as the measurement unit instead of occasions. The main argument is that the concept of occasion is subject to different cultural interpretations.

5.3 Two surveys of surveys

The EUROHIS survey of surveys proved to be a very useful source in studying the variation in methods that were used to record alcohol consumption in HIS. While most of the HIS contained questions on alcohol consumption, the questions differed widely in scope, detail, definitions of concepts and units of measurement (Table 5.1).

Table 5.1. Variation among HIS in questions on alcohol consumption

	No. of HIS
<i>A. Prevalence of drinking</i>	
Use vs. abstinence	
Unspecified	5
One month	1
One year	3
Lifetime	2
Never	3
Beverage-specific use	5
<i>B. Frequency of drinking</i>	
Overall drinking frequency	
Unspecified	4
One week	3
One month	5
One year	3
Beverage-specific	3
<i>C. Volume of alcohol consumption</i>	
Typical or average volumes	
Per occasion (cf. binge)	5
Per day	4
Per week	4
Per month	3
Per year	3
Beverage-specific	5
Actual volumes	
Per day	3
Per week	4
Per month	3
Beverage-specific	5
<i>D. Binge drinking</i>	
“Objective” criteria: Typical frequencies of drinking 4+, 5+ etc. per occasion	1
“Subjective” criteria: Drinking so much as to feel the effects	4
<i>E. Drinking in specific contexts</i>	2
Typical or actual drinking: alone, at meals, in pubs or bars, at sports events, before driving	
<i>F. Other aspects of drinking</i>	10

In another review of surveys, ECAS (Norström 2002) summarized the data that were collected on drinking patterns in general population surveys in the EU and Norway from 1950 to 1995 (Simpura & Karlsson 2001a). The materials were collected by a network of national experts in the EU countries. Table 5.2 shows the availability of data on alcohol in

surveys in the 1990s, as well as the extent of coverage of different aspects of alcohol consumption. It should be noted that the figures in the cells are based on the expert opinion of the authors, and cannot be directly found in the original country reports.

These two surveys of surveys have shown that, in the 1990s, the overall availability of survey data on drinking patterns has been best in the Nordic countries, which is no doubt influenced by the keen interest in alcohol control policies in these countries. The Netherlands and later the United Kingdom have also frequently collected a broad range of data on drinking patterns. Recent Italian studies have also added to the availability of data. In these leading countries, the main source of information comes from special studies on drinking patterns, except for the United Kingdom where data on drinking patterns are included in the general health surveys.

In the majority of the other countries that were reviewed, general health surveys are the most important source of alcohol data. Occasional special studies on alcohol have been conducted in Austria and Germany, while in some northern European countries and in Italy during the 1990s, data on trends are available from series of special surveys on drinking patterns. Finally, there are a number of countries where trend data on drinking patterns is virtually nonexistent.

The countries that collect the most data also have the broadest coverage of aspects of alcohol consumption in their surveys. In contrast, the countries that have few survey data generally include only a few different aspects of drinking. The standard question relates to drinking vs. abstinence, often as part of a question on overall drinking frequency. Other aspects of drinking are covered far less frequently, with the result that it would be difficult to arrive at sensible international comparisons, even on a cross-sectional basis. Comparisons of trends in drinking patterns would not be possible at all with the currently collected data.

Table 5.2. Availability and coverage of data collected on alcohol in surveys in the EU member countries and Norway in the 1990s**Legend***Availability of data*1 Some
2 A fair amount
3 Plenty*Coverage of key aspects of drinking behaviour (A to F)*0 Not covered
1 Sometimes covered
2 Covered in about half of the studies
3 Covered in all or most studies*Countries in the order presented in the table*

AUT	Austria	BEL	Belgium	DEU	Germany	DEN	Denmark
IRE	Ireland	NET	Netherlands	UNK	United Kingdom	SPA	Spain
FRA	France	GRE	Greece	ITA	Italy	POR	Portugal
FIN	Finland	NOR	Norway	SWE	Sweden		

	Beer-drinking countries							Wine-drinking countries					Former spirits-drinking countries		
	A U T	B E L	D E U	D E N	I R E	N E T	U N K	S W E	F R A	G E R	I T A	P O R	F R A	N O R	S W E
Availability of data	1	1	1	1	2	2	3	1	2	1	3	1	3	3	3
<i>Coverage</i>															
A. Prevalence of drinking	3	3	2	3	3	3	3	2	2	1	2	1	3	3	3
B. Frequency of drinking	0	1	2	1	1	1	2	1	2	1	3	2	3	3	3
C. Volume of alcohol consumption	2	2	1	2	1	1	3	1	1	1	1	1	3	3	3
D. Binge drinking	1	2	2	1	0	1	2	1	2	2	1	0	3	1	1
E. Drinking in specific contexts	1	1	0	2	0	2	1	1	2	1	3	2	3	1	1
F. Other aspects of drinking	1	0	1	2	0	2	2	1	1	0	2	1	2	2	3

5.4 Protocol analysis: cognitive processing and cultural interpretation of questions

The 1999 draft common instrument was used in protocol analysis exercises in 2000 in Finland, Germany, Italy and the Netherlands. It was originally intended also to collect French material, but these attempts met various obstacles. The basic idea of the protocol analysis approach is that, even if questionnaires are technically identical in different countries, it is possible that the questions are interpreted differently. These different interpretations could arise from different cultural attitudes to drinking or from cultural differences in the cognitive strategies that are used to respond to survey questions (Midanik & Hines 1991).

The findings from the protocol analysis exercises cannot be easily quantified, as the data are basically qualitative in nature. The protocol analysis interviews were recorded and the responses were transposed into a written text. A quantitative code was then developed so that this written text could be further analysed. The results suggest that variation across countries should be accounted for, as related both to the use of cognitive strategies and to the cultural sensitivity of the survey questions on drinking. Some of the differences

between the countries resulted from cultural differences in answering the questions, rather than from differences in the actual behaviour studied. For example, the social setting of meals can vary enormously between countries; the evening meal may be an instant snack in the north of Europe, but may last several hours in more southern European countries (see Raitasalo et al. 2002 for further details). These variations may be relatively small, however, compared to the differences in the accuracy of reporting of alcohol consumption, to be discussed in the next section. The overall effect is somewhat similar to the variation that is related to small sample size. Nevertheless, the differences between countries need to be interpreted cautiously, and only very large differences should be accepted as significant.

5.5 The ECAS 2000: a huge variation in underreporting

Underreporting of alcohol consumption is a pertinent problem in drinking habit surveys all over the world. Survey-based estimates of per capita alcohol consumption typically range from between 40% and 60% of the actual, statistically recorded or estimated level of consumption (Simpura 1996). People forget or deliberately produce understatements. Underreporting can be different within one country from one year to another and it may be different for different beverage types (see, for example, the detailed report on the Finnish experience by Mustonen et al. 1999).

These variations in reporting rates can be checked in national surveys, but other aspects are more intractable. The validity of reporting of alcohol consumption may vary between various socio-demographic groups. It may also be different for different components of drinking, such as whether one uses alcohol at all (abstinence/prevalence of use), frequency of drinking and intake per occasion, etc.

In studies where no attempt is made to estimate the volume of the respondent's alcohol consumption, or where this is impossible (e.g. the European School Survey Project on Alcohol and Drugs (ESPAD), see Hibell et al. 2001), there is no way of knowing the extent to which differences between countries are actually differences in reporting rather than differences in behaviour. Thus, there is a risk of self-deception in health behaviour surveys: if the researcher does not compare his or her findings with independent parallel indicators, or if questions that would make such comparisons possible are not included, there is a risk of overlooking important aspects of comparability.

In May 2000 ECAS conducted a six-country comparative study on drinking patterns. Telephone interviews were conducted with 1000 respondents over 18 years of age in six countries (Finland, France, Germany, Italy, Sweden and the United Kingdom) (see Leifman 2001 for further details). The questionnaire followed largely the same structure as the EUROHIS draft instrument, but had many differences in wording. The questionnaire was carefully translated by experts, and data collection followed a standardized procedure in all countries. It was found that completeness of reporting differed widely. In France and Germany the survey-based consumption estimate was below 30% of the actual consumption; in Finland and Sweden it was between 40% and 50%; in Italy about 55%; and in the United Kingdom an unbelievable 90%. The figure for the United Kingdom was unprecedented in the history of alcohol research. The researchers checked all possible technical faults that could explain such an exceptionally high figure, but nothing was found.⁴

⁴ To "adjust" the data for international comparison, it is tempting to use the inverse of the coverage rate as a calibration, or the correction coefficient. For example, if the reporting rate is 0.4, then consumption estimates should be multiplied by 2.5 and frequency estimates by 1.58 (as the square root of 2.5, on the assumption that underreporting is evenly distributed between frequencies of drinking and intake per occasion; there is a

The lesson from this experience is that comparative data from HIS between countries or over time cannot be taken at face value. The possibility of different rates of underreporting (or indeed overreporting) should be considered and discussed. Furthermore, this is probably not a peculiarity of alcohol research but a general problem in studying health-related behaviour.

5.6 EUROHIS field-testing in 2001

The EUROHIS field-testing that was conducted in France, Germany and the United Kingdom in 2001 used a slightly modified version of the instrument that had initially been proposed. The modifications were purely stylistic, and did not change either the structure or the components of the instrument. The field-test questionnaire (see Appendix 5.1) covered all the basic components of the composite indicator “alcohol consumption” described earlier in this chapter:

- abstinence (with two reference periods, of one year and one month);
- overall drinking frequency (number of drinking days in the last four weeks);
- average alcohol intake per drinking day constructed from:
 - beverage-specific drinking frequencies for four beverage types; and
 - beverage-specific intake per day when the beverage was consumed;
- frequency of occasions with high intake, both objectively (five drinks or more) and subjectively (“enough to feel it”); and
- frequency of drinking into six specific drinking situations (circumstances).

The questions on drinking frequency and intake per drinking day provided the basis for constructing a measure of the volume of alcohol intake during the four-week period preceding the interview.

The performance of the field-test questionnaire was assessed by looking at the following five aspects:

1. The findings from the qualitative observations were compared with those for the other EUROHIS instruments, and the level of non-response was investigated.
2. The field-test results for the basic components of alcohol consumption were compared with findings from other recent studies in France, Germany and the United Kingdom, as described in the compilation by Simpura & Karlsson (2001a,b) and in a six-country study conducted in 2000 and reported by Leifman (in press).
3. The components within the alcohol instrument were checked for internal consistency.
4. Further alcohol-specific indicators were constructed to compare the EUROHIS findings with those from other studies.
5. An overall assessment of the instrument’s performance was made.

5.6.1 Qualitative observations and non-response

The qualitative information gathered as part of the field-testing included a report for each question as to how often the respondents needed clarification and repetition, interrupted the

multiplicative effect here, where Volume = Mean frequency × Mean intake per occasion). However, such simple correction procedures will always be arbitrary.

interviewer while he or she was reading answer categories, expressed doubt, or had problems with the answer categories. The qualitative evaluation therefore reflects the overall feasibility of the questions.

The qualitative results were analysed from three different angles. First, the reaction to the set of alcohol questions was compared with the reactions to the other EUROHIS instruments. This indicated that the alcohol questions were neither more difficult nor easier than the other sets, at least not in any systematic way (data not presented here). Second, a comparison between the individual alcohol questions showed that in France and Germany there were few differences in the need for clarification, while in the United Kingdom the questions that referred to daily volume of alcohol consumption were more difficult to answer than the other questions. Third, a comparison between the three countries revealed very few differences between France and Germany, whereas the United Kingdom respondents had consistently greater difficulty in answering. The proportion of French and German respondents who had difficulties ranged from 2% to 24% and from 4% to 16%, respectively, with the easiest question being the frequency of drinking “enough to feel it”, and the most difficult question being daily intake of “other alcoholic beverages”. The proportion of United Kingdom respondents who had difficulties was generally higher, at 13–33%. The easiest question here was that on frequency of drinking spirits, and the most difficult question was that on daily intake of spirits.

As a rule of thumb, a question may be problematic if more than 20% of the respondents have difficulties in answering it. From this perspective, one question in France, no single question in Germany but seven questions in the United Kingdom were too difficult for the respondents. In particular, the United Kingdom respondents had difficulty in answering the questions that related to beer drinking, the quantity of spirits consumed, and the “objective” frequency of intoxication. These findings from the qualitative analysis suggest that cultural variation may play a role in the differences between the results from different countries, even when technically and linguistically identical questions on alcohol consumption are used.

Concerning non-response rates, a little under 10% of interviews had so many missing data, “don’t know” answers or refusals that it was impossible to calculate beverage-specific alcohol consumption figures for these individuals. This is a significant proportion of the total sample and may introduce a bias in the analysis.

5.6.2 Comparison of findings on basic alcohol components with those from other studies

A rough comparison was made with other alcohol studies, in order to detect whether the findings from the field-test differed radically from expected values. The results from the literature are given here as ranges (based on the results reported by Simpura & Karlsson 2001a and in some cases by Leifman, in press). Where results for the age bracket of the field study (i.e. 35–75 years) were not available, these ranges were based on expert opinion.

With respect to the prevalence of non-drinking, or the abstinence rates (Table 5.3), the results of the field-test revealed relatively low abstinence rates for men and relatively high abstinence rates for women. With respect to overall drinking frequencies (Table 5.4), the findings revealed expected values for men but relatively high frequencies for women. However, the question that was used in the field-test (the number of drinking days in the last four weeks) was different from the questions that have usually been used in surveys on drinking patterns (typically: “How often do you drink ...”, where the answer categories are daily, almost daily, etc.).

Table 5.3. Non-drinking rates: comparison of field-test data with other studies (percentages of respondents between 35 and 75 years of age)

		France	United Kingdom	Germany
Field-test	Men	4	8	10
	Women	25	20	19
Typical ranges	Men	5–10	7–15	10–15
	Women	15–25	12–20	15–20

Note: Non-drinkers are those who reported no consumption of alcohol during the 12 months before the interview.

Table 5.4. Mean number of drinking days in the last four weeks: comparison of field-test data with other studies

		France	United Kingdom	Germany
Field-test	Men	18	15	14
	Women	13	11	10
Typical ranges	Men	12–20	10–15	9–15
	Women	5–10	5–10	5–10

With respect to the prevalence of occasions with high alcohol intake (Table 5.5), the reference time unit in the field-test differed from that used in most other studies. As expected therefore, the rates in the field-test were higher than those reported in earlier studies. It was also found, as expected, that the United Kingdom “intoxication rates” were higher than those in France and Germany.

Table 5.5. Number of days/occasions with five drinks or more in the last four weeks: comparison of field-test data with other studies

		France	United Kingdom	Germany
Field-test	Men	10	48	17
	Women	12	20	6
Typical ranges	Men	8–12	30–40	8–15
	Women	2–3	3–4	3–4

Note: In the field-test, the question enquired about the number of days with five drinks or more during the last four weeks. In most other studies, the question is how often the respondent drinks five drinks or more, with response alternatives of daily, almost daily, etc.

5.6.3 Internal consistency

Those aspects of alcohol consumption that were expected to correlate or to show parallel trends were investigated. There were three such indicator components in the field-test: frequency of drinking, annual alcohol consumption per individual, and frequency of occasions with high alcohol intake.

Two frequency measures were available (Table 5.6): an overall drinking frequency and a sum of the beverage-specific drinking frequencies (beer, wine, spirits and other alcoholic beverages). The summed frequency is usually higher than the overall one, as the latter also contains occasions where several beverage types are consumed simultaneously. This was confirmed by the field-test data in all countries and for both genders.

Table 5.6. Comparison between two frequency measures (overall: based on one question on overall frequency; sum: based on the sum of beverage-specific frequencies). Figures are the number of drinking days in the last four weeks

		France	United Kingdom	Germany
Overall	Men	18	15	14
	Women	13	11	10
Sum	Men	24	19	18
	Women	16	14	13

Two parallel measures of annual alcohol intake were derived from the questions on beverage-specific drinking frequencies and volumes per day, and the related single questions (how many drinking days in the last four weeks; how many drinks per drinking day). The correlations between the two measures were expectedly high, and the three countries showed similar patterns.

With respect to prevalence of high alcohol intake, the “subjective” frequency of drinking (enough to feel the effect) was compared with the “objective” frequency of drinking (five drinks or more in a day). It could be expected that not all days with five drinks or more contain occasions where the respondent drank enough to feel the effect. Thus the objective measures should give a higher prevalence of high alcohol intake than the subjective ones. This appeared to be true in most cases in the field-testing data. French men scored surprisingly lower on both indicators than French women. This could indicate some problems in the cultural interpretation of the questions in France. Otherwise, all differences between men and women and between countries were similar for both indicators.

Table 5.7. Comparison between “objective” and “subjective” measures of high alcohol intake in the field-test. Figures are the mean number of days in the last four weeks

		France	United Kingdom	Germany
Objective	Men	0.8	4.0	1.4
	Women	1.1	1.7	0.5
Subjective	Men	0.1	3.0	1.2
	Women	0.5	1.8	0.3

Note: “Subjective” refers to “drank enough to feel it”; “objective” refers to “drank 5 drinks or more”.

5.6.4 Construction of further alcohol indicators for comparison with other studies

Further “derived indicators” were calculated using the replies to several of the field-test questions. These derived indicators related to the annual per capita alcohol consumption (Table 5.8) and, as a further derivative of this, the male : female ratio of alcohol consumed (Table 5.9). Respondents in the EUROHIS field-test were aged between 35 and 75 years, while the sales figures for alcohol consumption related to the entire population of 15+ years. This produces differential effects on the comparability of the consumption figures from the two sources in different countries, and for men and women. Many of the United Kingdom studies have revealed that younger drinkers (below 30 years of age) drink more than their elders, while in France high annual alcohol intake is more typical in older male groups. It is also likely that older German respondents drink slightly more than the younger ones (see, for example, Simpura & Karlsson 2001a,b).

In Table 5.8 the field-test findings are compared with estimated per capita alcohol consumption figures in the population aged 15 years or more, based on 1999 health for all data from the WHO Regional Office (www.euro.who.int/HFADB). The health for all data are based on sales statistics and do not take into account unrecorded consumption (e.g.

home-brewed alcohol, etc.), so the consumption figures ideally should not be reported as single figures but rather expressed as ranges. The data comparison in Table 5.8 is further influenced by the possibility that underreporting may vary widely from one country to another, as discussed earlier in this chapter.

Table 5.8. Volume of consumption (litres of pure alcohol per head): comparison of EUROHIS field-test data with officially registered sales figures

Alcohol consumption per capita (15+ years)	France	United Kingdom	Germany
EUROHIS field-test	10.2	9.1	10.4
Health for all data 1999	13.3	10	12.6

On the assumption that the EUROHIS field-test data are valid, and given the above-mentioned differences between countries in the associations of age and drinking, the underreporting in the United Kingdom would seem to be much smaller than in France and Germany. This result is similar to the findings of Leifman (2001) discussed earlier in this chapter. The important observation is that underreporting is significant, as are the differences in underreporting between countries.

The field-test data also indicated beverage-specific consumption rates that corresponded with those from other sources. As expected, wine was the overwhelmingly dominating beverage in France, while beer dominated in Germany and the United Kingdom.

The final validity check of the field-test data related to gender differences in overall alcohol consumption. Here again, age effects may be different for men and women in different countries. In European general population surveys, the share of alcohol consumed by women is about 30%. The figures in Table 5.9 are therefore roughly in the expected range for each country. The male : female ratio of 3 means that women drink one-quarter of all the alcohol consumed.

Table 5.9. Male/female ratios of alcohol consumption per capita: comparison of EUROHIS field-test with data from other studies

	France	United Kingdom	Germany
EUROHIS field-test	3.6	3.1	2.6
Typical ranges	3.0–3.5	2.5–3.5	2.5–3.0

5.6.5 Overall assessment of the performance of the EUROHIS alcohol instrument

The field-test results did not reveal any systematic deviations from the findings that would be expected on the basis of existing research. A few unexpected findings that could not be explained were encountered, but these were of minor importance. It was therefore concluded that the instrument worked well in western Europe and did not require any major restructuring or improvement. Nevertheless, instructions for interviewers and respondents will be required, especially for the questions that relate to the amount of alcohol consumed. Finally, despite the good overall fit of the findings with results from earlier studies, the field-test reminds us that survey findings should not be used at face value in international comparisons. There is always culturally determined variation in HIS responses, even when equivalently translated and technically identical questions are used.

5.7 Recommended common survey instrument for alcohol consumption

It is often practical that a questionnaire on alcohol consumption contains a number of questions that are relatively unimportant in themselves. It is thus reasonable to begin by asking whether the respondent drinks at all, although prevalence of abstinence is in itself of secondary importance. Similarly, it is technically useful to ask about various beverage types, although there is very little evidence of beverage-specific health effects. Asking questions separately on each beverage type helps to estimate the per capita alcohol consumption, which again is necessary information when assessing the validity of the data. An alternative would be to ask about the number of drinks (units, glasses, etc.) per day, and separately for weekdays and weekends. This would also provide an estimate of per capita consumption in the sample, and important context-specific information on drinking on working days vs. drinking at weekends. Such an alternative set of questions has therefore been included at the end of the recommended instrument.

The most difficult problem when using survey data on alcohol consumption in international comparisons is underreporting, which varies from one country to another, from one beverage type to another, and from one point of time to another. This cannot be remedied by improving survey instruments, and is a factor that must be given serious consideration.

There are other problems in proposing a common all-European instrument to be used in HIS.

- *Abstinence/prevalence of use.* There is cultural variation in the definition of an alcoholic beverage, hence the need to ask about drinking of different beverage types – with the same beverage type probably having different characteristics in different countries. In some countries, for example, the word “alcohol” refers only to distilled beverages. In many other countries, respondents do not always recognize mild beverages (beer, cider) as being alcoholic.
- *Frequencies/volumes.* It is difficult to identify scientifically relevant and culturally understandable time frames for questions about alcohol. The alternatives range from one day to one year or even a lifetime. In countries where daily drinking is common, the time frame should be short, whereas a longer period would be appropriate in countries with lower drinking frequencies.
- *Volumes/binge drinking.* It is not easy to determine comparable measures (units, sizes of glasses, etc.) for alcoholic beverages. While it is important to use units that are understandable by the general population in each country (e.g. the number of glasses or bottles), these should also allow further calculations in common units, such as centilitres of alcohol, to be suitable for cross-country comparisons.
- *Binge drinking/contexts.* There is also cultural variation in the definition of a drinking occasion in terms of time, company, place and context, and the concept of “binge drinking” varies between countries. For some people, having a drink at a party may only be an episode within something that they would not call a drinking occasion, whereas others may regard the whole party as a drinking occasion.

The recommended instrument presented in Appendix 5.1 was designed to deal with these problems, as well as to provide data that should be comparable with the mainstream of existing survey data on drinking patterns. It consists of two parts: one core element that should always be included, and an optional element that can be applied when there are resources for additional data collection.

The EUROHIS instrument has also been compared with the set of alcohol questions included in the WHO Survey on Health and Health System Responsiveness (WHO 2000). Of the 14 questions in this survey that concern alcohol, four aim at describing drinking patterns and alcohol consumption, while the others are directed at symptoms of alcohol dependency and alcohol-related diseases. Of these four questions, two (overall drinking frequency and binge drinking) are similar to the respective EUROHIS questions. The other two (lifetime abstinence and typical intake per occasion) are not included in the EUROHIS proposal because they are too crude indicators of drinking in the European context. For example, the expression “How many standard drinks...” does not fit well within a European context. Slight modifications were made in the EUROHIS alcohol instrument, however, to make it more comparable with the WHO questionnaire (WHO 2000).

5.8 Conclusions

The task of developing an all-European survey instrument on alcohol consumption has been quite difficult, mostly owing to problems that cannot be remedied by the use of standard survey methods, such as the use of identical techniques for sampling and data collection and identical cross-translated questionnaires. The use of standard methods cannot guarantee that the results will be comparable. Particular attention is required to deal with cultural variation in responses to survey questions. Even the most sincere people provide biased reports on their drinking and other health-related behaviour. This bias also varies according to time and place.

With these caveats, the EUROHIS alcohol consumption network proposes here a common instrument for use in European HIS. This instrument aims to cover four basic health-related elements of drinking: overall prevalence, volume of consumption (based on recording typical drinking frequencies and typical intake per occasion), binge drinking (i.e. prevalence of days of high intake) and prevalence of drinking in certain risk contexts. The recommended instrument also allows modifications in order to meet specific cultural requirements related to beverage types and units of measurement, such as sizes of bottles and glasses.

References

- DAWSON, D. & ROOM, R. (2000) Towards agreement on ways to measure and report drinking patterns and alcohol-related problems in adult general population surveys: the Skarpö conference overview. *Journal of substance abuse*, **12**: 1–23.
- EDWARDS, G. ET AL. (1994) *Alcohol policy and the public good*. Oxford, Oxford University Press.
- HIBELL, B. ET AL. (2001) *The 1999 Espad Report. The European School Survey Project on Alcohol and Drugs. Alcohol and other drug use among students in 30 European countries*. Stockholm, CAN.
- KNIBBE, R. & BLOOMFIELD, K. (2001) Alcohol consumption estimates in surveys in Europe: comparability and sensitivity for gender differences. *Substance abuse*, **22**: 23–47.
- LEIFMAN, H. (2001) Homogenisation in alcohol consumption in the European Union. *Nordic studies on alcohol and drugs*, **18**(English Supplement): 15–30.
- LEIFMAN, H. (in press) A comparative analysis of drinking habits in six European countries year 2000. *Journal of drug issues*.
- MIDANIK, L.T. & HINES, A.M. (1991) “Unstandard” ways of answering standard questions: protocol analysis

- in alcohol survey research. *Drug and alcohol dependence*, **27**: 245–252.
- MUSTONEN, H. ET AL. (1999) *Finnish drinking habits in 1968, 1976, 1984, 1992 and 1996*. Helsinki, STAKES (Themes 7/1999).
- NORSTRÖM, T. (2001) *Per capita alcohol consumption and all-cause mortality in 14 European countries*. *Addiction*, **96**(Suppl. 1): S113–S118.
- NORSTRÖM, T., ED. (2002) *Alcohol in postwar Europe. Consumption, drinking patterns, consequences and policy responses in 15 European countries*. Stockholm, Almqvist & Wicksell.
- RAITASALO, K. ET AL. (2002) *Behind questions: intercultural differences in answering questions on drinking behavior in Health Interview Surveys – A protocol analysis study in four European countries*. Helsinki, STAKES (Themes from Finland 2/2002).
- REHM, J. (1999) Draining the ocean to prevent shark attacks? *Nordic studies on alcohol and drugs*, **16** (English Supplement): 46–54.
- REHM, J. & GMEL, G. (2000) Aggregating dimensions of alcohol consumption to predict medical and social consequences. *Journal of substance abuse*, **12**: 155–169.
- ROOM, R. (2000) Measuring drinking patterns: the experience of the last half century. *Journal of substance abuse*, **12**: 24–33.
- SIMPURA, J. (1996) Trends in alcohol consumption and drinking patterns: lessons from world-wide development. In: Holder, H. & Edwards, G., ed. *Alcohol and public policy. Evidence and issues*. Oxford, Oxford University Press, pp. 7–37.
- SIMPURA, J. & KARLSSON, T. (2001a) Trends in drinking patterns in 15 European countries, 1950 to 1995. A review of existing research. *Nordic studies on alcohol and drugs*, **18**(English Supplement): 31–53.
- SIMPURA, J. & KARLSSON, T. (2001b) *Trends in drinking patterns in fifteen European countries, 1950 to 2000. A collection of country reports*. Helsinki, STAKES.
- WORLD HEALTH ORGANIZATION (2000) *International guide for monitoring alcohol consumption and related harm*. Geneva, World Health Organization (document WHO/MSD/MSB/00.4).

Appendix 5.1

EUROHIS recommended common instrument for alcohol consumption

NB. CORE QUESTIONS ARE MARKED WITH AN ASTERISK (*)

Prevalence of use/abstinence

*1a. During the past 12 months, did you drink any alcohol (beer, wine, spirits, [other local beverages])?

Yes/No

If “No”, finish.

1b. Did you drink any alcohol (beer, wine, spirits, [other local beverages]) in the past 4 weeks?

Yes/No

If “No”, finish.

Frequency of drinking

2a. In the past 4 weeks, on how many days did you drink any beer?

2b. In the past 4 weeks, on how many days did you drink any wine?

2c. In the past 4 weeks, on how many days did you drink any spirits?

2d. In the past 4 weeks, on how many days did you drink any [other local beverages]?

* 2e. (To sum up,) On how many days in the past 4 weeks did you drink any alcohol?

For each question 2a–2e: On ___ days (number of respective drinking days per month)

Don't know

Intake per occasion

3a. On a day when you drink beer, how much do you usually drink? You may answer in [different units, modifiable in the countries]

Bottles of 70–75 cl ___ (number) in a day

or

Glasses of 12–20 cl ___ (number) in a day

3b. On a day when you drink wine, how much do you usually drink? You may answer in [different units, modifiable in the countries]

[two or three alternatives] ___ (number) in a day

3c. On a day when you drink spirits, how much do you usually drink? You may answer in [different units, modifiable in the countries]

[two or three alternatives] ___ (number) in a day

- 3d. On a day when you drink [other local beverages], how much do you usually drink? You may answer in [different units, modifiable in the countries] [two or three alternatives] ____ (number) in a day

NB. THE SETS OF QUESTIONS 2 AND 3 ALLOW THE CALCULATION OF A QUANTITY–FREQUENCY MEASURE (QF) OF ALCOHOL CONSUMPTION FOR EACH RESPONDENT.

- *3e. (To sum up,) On a day when you drink any alcohol, how much do you usually drink altogether? Sum up all drinks, bottles of beer, glasses of wine, measures of spirits, etc.
- 1–2
 - 3–4
 - 5–6
 - 7–9
 - 10 or more

Binge drinking

- *4a. In the past 4 weeks, on how many days did you have at least one bottle of wine and/or 20 cl of spirits (approximately 5 drinks) and/or 5 bottles [or equivalent amount in local units, 1 bottle = 30–40 cl] beer in one drinking occasion?
____ Number of days with 5+
- 4b. In the past 4 weeks, on how many days did it happen that you drank “enough to feel it” [or respective local culturally valid expression]?
____ Number of days with “enough to feel it”

Drinking in specific contexts

During the past 7 days, on how many days did you drink alcohol in the following contexts:

- 5a. Alone
- 5b. At a bar, pub, tavern or café
- 5c. At a sports or entertainment event
- 5d. At lunch
- 5e. At dinner
- 5f. Before driving a car

Response: [Number of days, 0–7; 9 = Don’t know]

.....

Alternative approach for questions 3a–3c (surveying weekends and working days separately):

(These items were not field-tested in 2001)

- 6a. On how many of the weekly five working days do you usually drink alcohol?
____ Number of days (0–5)

- 6b. How many (drinks/units/glasses) do you typically drink on such a working day when you drink alcohol?
___ Number of (drinks/units/glasses)
- 6c. Do you usually drink alcohol on Saturday or Sunday or both?
- 1 No
 - 2 Yes, on Saturday only
 - 3 Yes, on Sunday only
 - 4 Yes, on both days
- 6d. How many (drinks/units/glasses) do you typically drink on such a weekend day when you drink alcohol?
___ Number of (drinks/units/glasses)

Chapter 6.

Development of a common instrument for physical activity

Pekka Oja

Health Promotion Research, UKK Institute, Tampere, Finland

6.1 Introduction

The scientific evidence that has accumulated during the 1990s (for example, Bouchard et al. 1990, 1994; Kesäniemi et al. 2001; US Department of Health and Human Services 1996) indicates consistently that physical activity substantially reduces the risk of coronary heart disease, stroke, obesity, type II diabetes, hypertension, colonic cancer and osteoporotic fractures among middle-aged and older adults. Moreover, physical activity benefits growth and development in children and young adults and helps to maintain functional capacity and independence in ageing people. In view of these potential health benefits and the high prevalence of physical inactivity in most countries, the public health impact of greater physical activity can be very significant.

A new understanding of the characteristics of health-enhancing physical activity (HEPA) has also developed. Half an hour a day of physical activity of moderate intensity provides many of the health benefits for an inactive population (US Department of Health and Human Services 1996). To survey the population levels and patterns of HEPA, therefore, a new survey instrument is needed that is capable of measuring the key characteristics of HEPA.

Simultaneously with the development of the EUROHIS project, an international initiative – IPAQ (International Physical Activity Questionnaire) – was started in order to develop standard interview methods for measuring HEPA in population surveys. The global IPAQ network proposed a set of physical activity questionnaires. These have formed the basis for the development of the EUROHIS instrument for physical activity.

6.2 Methodology

The development of the EUROHIS physical activity instrument followed six main stages. First, the concept of HEPA was discussed. This yielded a working definition of HEPA that was used as the conceptual framework throughout the work. Next, the findings from the EUROHIS “survey of surveys” were reviewed. This provided information on the instruments that were currently being used in European countries to measure physical activity. The third stage was the development of a draft instrument in connection with the global IPAQ project. This draft was then pre-tested in 2000 alongside other EUROHIS draft instruments. The reliability and validity of the proposed instrument was also assessed as part of the global IPAQ project. Last, the draft EUROHIS instrument was field-tested in 2001 as part of the overall EUROHIS questionnaire.

6.3 Definition of physical activity

Discussion at the first meeting of the physical activity network, in Berlin in April 1999, focused on defining the concept of physical activity. In line with the research evidence on HEPA, the following working definition was adopted for the development of the physical activity indicator:

Physical activity consists of bodily movements produced by the skeletal muscle contraction that increase energy expenditure above the basal level. Physical activity manifests itself in work or education, transport, domestic chores and recreation and its key characteristics are frequency, intensity, duration and continuation.

This definition makes it clear that not only leisure-time sports and exercise are included, but also common lifestyle activities that are carried out at work (e.g. physical labour), at home (e.g. housework, gardening) and as transport (e.g. walking, cycling).

The current general recommendation for health-enhancing physical activity calls for a total of half an hour of moderate intensity activity on most days of the week (US Department of Health and Human Services 1996). Information on the key characteristics of HEPA – duration, intensity and frequency – would allow an evaluation of the sufficiency of individuals' or populations' physical activity with respect to their health. As the IPAQ questionnaire was designed according to this concept, the IPAQ format was adopted for use in the EUROHIS project.

6.4 Survey of surveys

The EUROHIS survey of surveys included a section that enquired about the physical activity instruments currently being used in European national health surveys. Thirty-two physical activity questionnaires from 29 countries were received. Analysis of the 28 questionnaires that were provided in English revealed the following types of question:

- frequency of physical activity (15)
- self-classification of physical activity (8)
- quantitative description of physical activity (3)
- yes/no physical activity (1)
- amount of walking (1).

While about half of the surveys included questions on frequency of physical activity, these questions varied with regard to the type of physical activity asked about. Most surveys used an intensity scale to assess the degree of physical activity, but only a few considered the whole range of light/moderate/vigorous exercise. The survey methods also varied in terms of the activity domain asked about, i.e. whether the physical activity took place during leisure time, at work, in domestic work or as transport. The focus in the existing surveys was on physical activity undertaken during leisure time. Among all the questionnaires that were reviewed, only two physical activity instruments were identical.

The findings from the survey of surveys thus confirmed that there was little uniformity among the physical activity instruments currently used in Europe and, furthermore, that these did not conform to the IPAQ concept of physical activity. Although these data have not been further investigated as part of the EUROHIS project, further analysis of the survey of surveys data would provide important historical information on physical activity survey methodology in Europe.

6.5 Designing the IPAQ instrument

The IPAQ project was formally initiated in 1996 when a group of researchers proposed an international effort to develop a standard questionnaire for the purpose of public health surveillance of physical activity. In Geneva in 1998, supported by WHO and the US Centres for Disease Control and Prevention, and later by the Karolinska Institute in Sweden, physical activity researchers from 14 countries convened and started the development of questionnaires that could be used in population-based national surveys. The continuing international collaboration has drawn together most of the research centres in the world that currently work in the area of health-related physical activity assessment, and includes three members of the EUROHIS physical activity network. The choice of IPAQ meant that the EUROHIS project could benefit from the comprehensive, relevant and timely international work that was already in progress.

The IPAQ instrument aims to assess the current total amount of physical activity that is undertaken, by retrospectively asking about frequency, duration and intensity of all domains of physical activity, i.e. work/education, transport, domestic chores and recreation. The amount of time per week that is spent walking and sitting is also enquired about. By the end of 1999,¹ the IPAQ network had produced eight versions of the instrument, which varied with respect to length (short/long), reference period (usual week/last seven days) and method of administration (telephone/self-administered).

The proposed short IPAQ versions contain nine questions concerning activities at work, transport from place to place, indoor and outdoor domestic work, as well as exercise, recreation and sport during one week. Two questions deal with frequency and duration of all vigorous activity, two questions with frequency and duration of all moderate intensity activity, three questions with frequency, duration and intensity of walking, and two questions with time spent sitting.

The proposed long IPAQ versions contain 31 questions covering the intensity and duration of physical activity at work, during transportation, at home, and during leisure time. Two questions probing inactivity in terms of the amount of sitting are also included.

6.6 EUROHIS pre-testing: feasibility

In September/October 2000, as part of the EUROHIS project and together with draft instruments for the other proposed EUROHIS indicators, the IPAQ instrument (either the long or the short version) was pre-tested in Bulgaria, Hungary, Slovakia, Spain and Ukraine. There were 20–25 respondents in each country, most of whom were recruited from the survey institution's staff or from other conveniently available groups. Translations were produced without a formal protocol. The results were presented and discussed at the EUROHIS mid-term review in November 2000. The short physical activity questionnaire was considered too complex by Bulgaria and too non-specific by Hungary, while Slovakia, Spain and Ukraine reported that the long questionnaire was acceptable albeit somewhat difficult to complete.

¹ The global IPAQ project was completed in 2001. The experience has been reported as a research article (Craig et al. in press). The project will continue as an International Prevalence Study on Physical Activity (IPS), which will be coordinated by a research committee comprising experts in the fields of physical activity, public health and surveillance systems. For more details, see www.ipaq.ki.se.

6.7 IPAQ network field-testing: reliability and validity

During 2000, the IPAQ network compared the utility, feasibility, reliability and validity of the IPAQ instrument in six European countries: England (2 centres), Finland, Italy, the Netherlands, Portugal and Sweden. Each centre selected several of the eight questionnaire versions for testing. Convenience samples of between 30 and 200 respondents were used (see Table 6.1).

The test–retest reliability was evaluated by repeated administration of the same questionnaire within a maximum of eight days. Concurrent validity was assessed by comparing the short and long questionnaires that covered the same duration. Criterion validity was tested by comparing a selected short IPAQ questionnaire with the total count of a CSA accelerometer over the same period. The CSA accelerometer was fastened to the subject's belt and provided three-dimensional recordings of all body movements that generated acceleration beyond a set threshold. This method has been shown to accurately estimate total energy expenditure over the recorded time.

The Dutch and Finnish translations were produced following the standard IPAQ translation protocol and included the following steps: (a) translation from the original English version into the target language by two independent translators; (b) review of these translations by a group of bilingual people and production of a revised version; (c) translation of the revised version back into English by two different translators; and (d) review of these back translations and production of a final version by the group of bilingual people. The translations into Italian, Portuguese and Swedish were produced more informally.

Table 6.1 Methodology and sample characteristics in IPAQ field-testing, 2000

Centre	Bristol, England	Cambridge, England	Finland	Italy	Netherlands	Portugal	Sweden ^a
Gender	83 M 93 F	38 M 64 F	43 M 45 F	15 M 14 F	24 M 36 F	96 M 100 F	22/77 M 28/123 F
Age (years)	18–39	42–50	28–74	65–66	Mean 33	Mean 35	Mean 41 Mean 47
Test–retest reliability	S7S	SUT	SUS	S7S	S7S	SUS	SUS & S7T
Concurrent validity	S7S/L7S	–	S7S/L7S	–	–	SUS/LUS	SUS/LUS & S7T/L7T
Criterion validity	S7S/CSA	SUT & S7T/ heart rate, fitness	SUS/CSA, pedometer	S7S / fitness	S7S/CSA	–	S7S/CSA

^a Sweden had two samples.

IPAQ questionnaire versions

SUT: short/usual week/telephone; SUS: short/usual week/self-completed; S7T: short/past 7 days/telephone;

S7S: short/past 7 days/self-completed

LUT: long/usual week/telephone; LUS: long/usual week/self-completed; L7T: long/past 7 days/telephone;

L7S: long/past 7 days/self-completed

CSA: CSA accelerometer; Cambridge used long-term heart rate recording and cardio-respiratory fitness assessment as validation criteria, while Finland used the total number of steps as measured by a pedometer.

6.7.1 Results

The test–retest reliability coefficients (Table 6.2) for the short IPAQ questionnaires varied between fair and excellent, indicating in general an acceptable repeatability. There were no apparent differences between the questionnaire versions. Similar test–retest analysis of the S7S and SUS questionnaires in three centres outside Europe (Japan and the United States) yielded similar repeatability coefficients (0.73–0.89) (Craig et al. in press).

Criterion validity coefficients (Table 6.2) varied from 0.07 to 0.60. With the exception of the very low coefficient in the Swedish data, the results indicated results that were comparable to those of earlier studies on criterion validity of physical activity questionnaires (Kriska & Caspersen 1997). Results for criterion validity from non-European countries showed similar coefficients: 0.35–0.48 for the S7S questionnaire and 0.13–0.48 for the SUS questionnaire (Craig et al. in press).

Coefficients for concurrent validity between the short and long IPAQ questionnaires varied from 0.46 to 0.78. With the exception of Portugal, the agreement was acceptable. The non-European country results (Craig et al. in press) were similar, showing coefficients between 0.58 and 0.80, thus confirming the European observations.

Table 6.2. Reliability and validity of selected short IPAQ questionnaires

Country	N	Test–retest reliability ^a		Criterion validity ^b		Concurrent validity ^c	
		Questionnaire	ICC	Questionnaire	ICC	Questionnaire	ICC
England							
Bristol	149	S7S	0.69	S7S	0.58	S7S	0.74
Cambridge	101	SUT	0.85	–	–	–	–
Finland	87	SUS	0.65	S7S	0.60	S7S	0.72
Italy	29	S7S	0.68	S7S	0.28	–	–
Netherlands	60	S7S	0.95	S7S	0.37	–	–
Portugal	196	SUS	0.79	–	–	SUS	0.46
Sweden	50	SUS	0.70	S7S	0.07	SUS	0.78
	200	S7T	0.67	–	–	S7T	0.66

^a Test–retest reliability: ranked inter-correlation coefficient (ICC) between visit 1 and visit 2 for total weekly physical activity in MET minutes (multiples of resting metabolic rate, times the duration spent in the activity).

^b Criterion validity: ranked ICC between total weekly physical activity in MET minutes derived from the S7S questionnaire and total CSA counts.

^c Concurrent validity: ranked ICC between total weekly physical activity in MET minutes from the short and the respective long questionnaire.

S7S: short/past 7 days/self-completed; SUS: short/usual week/self-completed; S7T: short/past 7 days/telephone.

6.7.2 Conclusions

Some general conclusions could be drawn from the EUROHIS pre-testing and the IPAQ field-testing.

First, although the questionnaires were not considered easy to answer (e.g. problems were encountered with regard to estimating the time spent sitting, differentiating between vigorous and moderate intensity, identifying intensive home activities, and seasonal variation), the short questionnaire was considered acceptable for population surveys. The short version was especially favoured by managers of national health surveys, many of whom would not even consider the long version because it was considered too time-

consuming and complex. Some EUROHIS country representatives were hoping that a single physical activity question could be sufficient.

Second, although the reliability and validity properties of the “last seven days” and the “usual week” versions were similar, it was the general consensus of both the EUROHIS and the IPAQ networks that the “last seven days” version was preferable.

Third, the reliability and validity properties of the interview and self-completion versions were comparable. Countries may therefore choose between these in order to use data collection methods that they have used in the past or are going to use in their health surveys.

In view of the practical need to simplify the physical activity instrument, the following changes were made.

- The question on walking pace was deleted, as the difference between the total weekly MET minutes of walking with or without the walking pace was found to be negligible: 0.0–2.3% (and mostly less than 1%) in the European data.
- For measuring inactivity, only sitting on a weekday was included (i.e. not sitting on a weekend day). As the reliability properties of both the sitting questions were good and comparable, one question was considered sufficient.

6.8 EUROHIS field-testing 2001

On the basis of the studies described above, the short interview version of the IPAQ questionnaire that covered the last seven days (S7T) was used in the EUROHIS field-test in 2001. Computer-assisted telephone interviews were conducted in France, Germany and the United Kingdom. Four hundred interviews per country were administered, covering men and women aged between 35 and 75 years. Appendix 6.1 shows the physical activity instrument that was used for the interviews.

Within the framework of the field-test, it was considered appropriate to explore three main issues. First, the performance of the physical activity instrument as part of a comprehensive health interview survey was investigated by reviewing the level of completion for each question and the experiences of the interviewers. Second, the data were analysed to see whether there were differences between the field-testing countries. Third, a method for classifying the level of physical activity was investigated.

Because the field-test assessed the feasibility properties of the instrument as part of a global health survey, the results were considered unsuitable for descriptive or analytical epidemiological analysis.

6.8.1 Performance of the physical activity questions

Table 6.3 shows the completion rates for the physical activity questions by country and gender. The response rates were quite low, between 50% and 83%, for the questions that enquired about the number of days per week with vigorous and moderate activity, and women tended to respond less often than men. For the questions enquiring about the duration of moderate and vigorous activity per day, the United Kingdom completion rates were low (55–74%), while France and Germany largely showed full (100%) completion rates. The level of completion for the questions on walking and sitting was also high, generally over 90%. These results indicated that respondents had difficulties in comprehending “vigorous” and “moderate” intensity activities and in reporting the exact

number of days that they were undertaken. The reporting of the time spent doing these activities was also problematic in the United Kingdom.

Table 6.3. Level of completion rates for physical activity questions, by country and gender

Question	France						Germany						United Kingdom					
	Men (196)		Women (204)		All (400)		Men (184)		Women (217)		All (401)		Men (197)		Women (205)		All (4002)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Days with vigorous activity last week	115	58.7	102	50.0	217	54.2	119	64.7	122	56.2	241	60.1	125	63.5	112	54.6	237	59.0
Total time in minutes per day with vigorous activity	194	99.0	204	100	398	99.5	184	100	217	100	401	100	124	62.9	112	54.6	236	58.7
Days with moderate activity last week	125	63.8	130	63.7	255	63.7	144	78.3	179	82.5	324	80.8	146	74.1	124	60.0	270	67.2
Total time in minutes per day with moderate activity	196	100	204	100	400	100	184	100	217	100	401	100	145	73.6	123	60.0	268	66.7
Days with walking last week	184	93.9	191	93.6	375	93.7	161	87.5	204	94.0	365	91.0	181	91.9	189	92.2	370	92.0
Total time in minutes per day walking	196	100	204	100	400	100	184	100	217	100	401	100	180	91.4	189	92.2	369	91.8
Total time in minutes per day sitting	196	100	204	100	400	100	184	100	217	100	401	100	197	100	205	100	402	100

Feedback from the interviewers confirmed the difficulties that were indicated by the completion rates. The French experience revealed that people had difficulties in giving precise answers, but it was suggested that the provision of more examples would clarify the questions. (It is in fact suggested in the IPAQ instructions that culturally relevant examples should be used to illustrate the different intensity activities.) A specific suggestion was that, in the “sitting” question, sitting during meals should be included in the question. The German interviewers noted that the physical activity questions were in general long, so that the interviewers often had to repeat the questions. Furthermore, they pointed out that most of the respondents had difficulties in differentiating between vigorous and moderate physical activity. For example, housewives found it difficult to classify housework as vigorous or moderate. The German participants also observed that the inclusion of “sitting during meals” in the “sitting” question would be helpful.

6.8.2 Is physical activity different in the three western European countries?

The field-test results allowed some preliminary examination of possible country differences in physical activity. A consistent pattern emerged, with more of the French respondents reporting that they undertook vigorous (Table 6.4) and/or moderate (Table 6.5) physical activity every day of the week. Nevertheless, there were no marked differences in the total amount of time spent on these activities between the three countries. The walking data (Table 6.6) showed a similar pattern, with more French respondents reporting that they walked each day of the week. There were also more French respondents who walked more than two hours a day, when compared to the German and British respondents. Sitting for

more than six hours a day appeared to be more common in the United Kingdom than in France and Germany (Table 6.7).

These observations of different activity patterns in the field-test countries provide hypotheses for further studies using the IPAQ instrument.

Table 6.4. Percentage frequency and duration of vigorous physical activity, by country and gender

	France			Germany			United Kingdom		
	Men	Women	Total	Men	Women	Total	Men	Women	Total
<i>Number of days with vigorous activity last week</i>									
1	27	17	22	17	25	21	13	20	17
2	10	22	15	29	16	23	18	28	22
3	16	12	14	18	29	24	12	14	13
4	14	5	10	13	8	10	11	9	10
5	9	11	10	7	8	7	16	15	16
6	2	4	3	2	1	1	9	2	6
7	23	25	24	14	12	13	20	12	16
<i>Total time spent on vigorous activity per day</i>									
0–59 minutes	23	27	25	17	17	17	34	28	31
60–120 minutes	43	43	43	48	50	49	31	47	39
> 120 minutes	34	30	32	35	33	34	35	25	30

Table 6.5. Percentage frequency and duration of moderate physical activity, by country and gender

	France			Germany			United Kingdom		
	Men	Women	Total	Men	Women	Total	Men	Women	Total
<i>Number of days with moderate activity last week</i>									
1	14	19	17	16	12	14	10	12	17
2	15	16	16	17	23	20	30	22	22
3	12	17	15	19	15	16	14	14	13
4	5	4	4	6	7	7	13	5	10
5	4	8	6	10	11	11	13	17	16
6	7	2	4	3	2	3	1	3	6
7	41	29	35	29	30	29	17	26	16
<i>Total time spent on moderate activity per day</i>									
0–59 minutes	30	38	34	28	33	31	28	26	31
60–120 minutes	40	35	38	44	42	43	39	47	39
> 120 minutes	30	27	28	28	25	26	31	27	30

Table 6.6. Percentage frequency and duration of walking, by country and gender

	France			Germany			United Kingdom		
	Men	Women	Total	Men	Women	Total	Men	Women	Total
<i>Day with walking last week</i>									
1	2	3	2	5	2	3	7	5	6
2	5	5	5	9	7	8	9	6	7
3	5	7	6	6	7	6	10	8	9
4	2	4	3	9	6	7	6	5	5
5	8	1	5	5	11	8	6	13	9
6	2	1	2	6	5	5	8	6	7
7	76	77	77	61	61	61	55	58	56
<i>Total time per day walking</i>									
0–59 minutes	30	32	30	37	36	36	46	50	48
60–120 minutes	42	40	41	39	37	38	28	31	30
> 120 minutes	28	28	29	24	27	26	26	19	22

Table 6.7. Percentage frequency and duration of sitting, by country and gender

	France			Germany			United Kingdom		
	Men	Women	Total	Men	Women	Total	Men	Women	Total
<i>Total time per day sitting</i>									
0–59 minutes	5	6	5	2	0	1	0	5	2
60–120 minutes	12	17	15	15	14	14	8	7	7
121–180 minutes	15	21	18	11	16	14	10	20	15
181–240 minutes	22	14	18	17	22	20	14	18	16
241–300 minutes	11	14	13	17	15	16	14	8	11
301–360 minutes	7	10	8	14	11	12	11	11	11
> 360 minutes	28	18	24	25	21	22	44	32	35

6.8.3 Classification of physical activity

The ultimate objective of assessing health-related physical activity through national health surveys is to identify activity patterns that may indicate health risks within and between populations. As the IPAQ concept is still under development, there are no established criteria for health-related activity, such as the level of physical activity below which a disease risk increases significantly. These criteria can only be developed using new epidemiological research. In addition to the individual items in the IPAQ questionnaire, further physical activity measures can be derived from the responses. The IPAQ thus provides a quantitative estimate of the total weekly physical activity in MET minutes (MET

= multiples of resting metabolic rate), which is based on the weekly frequency and duration of vigorous and moderate physical activity and walking.

The total weekly MET minutes is calculated as follows:

1. Total weekly MET minutes for VIGOROUS activity
= weekly frequency (Q.1 in Appendix 6.1) \times daily duration (Q.2) \times 8 METs
[where 8 = generic MET estimate for vigorous activity]
2. Total weekly MET minutes for MODERATE activity
= weekly frequency (Q.3) \times daily duration (Q.4) \times 4 METs
3. Total weekly MET minutes for WALKING
= weekly frequency (Q.5) \times daily duration (Q.6) \times 3.3 METs
4. Grand total weekly MET minutes
= 1 + 2 + 3 from above.

Epidemiological evidence suggests that weekly physical activity that expends 1000 kcal is associated with up to 30% reduction in all-cause mortality rate, while activity that expends 500 kcal might have some favourable effect on all-cause mortality (Kesäniemi et al. 2001). As 1000 kcal per week is roughly equal to 1000 MET minutes, and 500 kcal roughly equal to 500 MET minutes, these values were tentatively set as classification criteria in the field-test data as follows:

- LOW level of physical activity: \leq 499 total MET minutes per week
- MODERATE level of physical activity: 500–999 total MET minutes per week
- HIGH level of physical activity: \geq 1000 total MET minutes per week

Using this classification, the field-test data yielded the following prevalence figures for level of physical activity: France 15% low, 7% moderate, 78% high; Germany 7% low, 8% moderate, 85% high; United Kingdom 16% low, 9% moderate, 76% high. All three countries appeared to have a similar high prevalence of high physical activity. Previous epidemiological assessment of physical activity has primarily taken into account only physical activity undertaken during leisure time, thus leading to greater prevalence of low physical activity. The advantage of the IPAQ is that it includes physical activity undertaken during occupational, transport and domestic activities. These observations highlight the need for new epidemiological research regarding the health-related criteria for physical activity, as measured for example by the IPAQ instrument.

6.9 Summary and conclusions

The main findings that emerged during the development of the EUROHIS physical activity instrument can be summarized as follows.

- Pre-testing of the feasibility of the questionnaire in five countries (Bulgaria, Hungary, Slovakia, Spain and Ukraine), using small convenient samples, indicated that the long version was acceptable albeit somewhat difficult to answer. The short version was considered complex and non-specific in two countries.
- Evaluation of the reliability and validity of the IPAQ instruments in six European countries indicated acceptable test–retest repeatability and criterion validity. The findings were similar to those for earlier physical activity instruments. Results from Japan and the United States were consistent with the European experience.

- On the basis of the measurement properties of the IPAQ instrument, the short questionnaire was considered suitable for population health surveys, as judged by experts in survey methodology and survey management. With regard to the recall reference period, it was recommended that the “last seven days” version be used. Both interview and self-completion forms were considered comparable and acceptable.
- The recommended EUROHIS physical activity instrument comprises seven questions: two each on vigorous and moderate intensity activity, two on walking and one on sitting. The interview version, which was used in the EUROHIS field-testing and which is consistent with the recommended IPAQ form, is given here in Appendix 6.1. (The IPAQ executive group has suggested two small changes in the wording of the questionnaire regarding the description of vigorous activity and the duration of different activities. All current IPAQ questionnaires are available at www.ipaq.ki.se.)
- The completion rates and the interviewer feedback in three European countries that field-tested the instrument indicated that respondents had some difficulties in identifying “vigorous” and “moderate” intensity activities, and in estimating the number of days and amount of time spent doing these activities. It is recommended that culturally and nationally relevant activity examples be added to assist the respondent in answering.
- The instrument provides data that are suitable for cross-national comparisons of health-related physical activity. Selected exploratory comparisons can serve as a basis for further studies.
- The proposed instrument can be used to estimate the total weekly activity in MET minutes for vigorous and moderate intensity activity and walking. Their summed total can be used as a measure of health-related physical activity, while the daily sitting time can be used as a measure of inactivity. However, it is not yet possible to propose a classification system for health-related physical activity. Further epidemiological research using the instrument is needed before reliable criteria for such classification can be established.

References

- AINSWORTH ET AL. (2000) Compendium of physical activities: an update of activity codes and MET intensities. *Medicine and science in sports and exercise*, 32(Suppl.): S498–S516.
- BOUCHARD, C. ET AL., ED. (1990) *Exercise, fitness and health. A consensus of current knowledge*. Champaign, IL, Human Kinetics.
- BOUCHARD, C. ET AL., ED. (1994) *Physical activity, fitness and health. International proceedings and consensus statement*. Champaign, IL, Human Kinetics.
- CRAIG, C.L. ET AL. (in press) The International Physical Activity Questionnaire (IPAQ): 12-country reliability and validity. *Medicine and science in sports and exercise*.
- KESÄNIEMI, Y.A. ET AL. (2001) Dose–response issues concerning physical activity and health: an evidence-based symposium. *Medicine and science in sports and exercise*, 33(Suppl.): S351–S358.
- KRISKA, A.M. & CASPERSEN, C.J. (1997) A collection of physical activity questionnaires for health-related research. *Medicine and science in sports and exercise*, 29(Suppl.): S1–S205.
- US DEPARTMENT OF HEALTH AND HUMAN SERVICES (1996) *Physical activity and health: a report of the Surgeon General*. Pittsburg, PA, President’s Council on Physical Fitness and Sports.

Appendix 6.1

EUROHIS recommended common instrument for physical activity (slightly modified after field-testing in 2001)

International Physical Activity Questionnaire [IPAQ] for young and middle-aged adults:

Short version covering the last seven days – telephone version

We are interested in finding out about the kinds of physical activity that people do as part of their everyday lives. I am going to ask you about the time you spent being physically active in the last seven days. Please answer each question, even if you do not consider yourself to be an active person. I will be asking you about activities you do at work, to get from place to place, as part of your house and yard work, and in your spare time for recreation, exercise or sport.

Think about all the vigorous activities that take **hard physical effort** that you did in the last seven days. Vigorous activities make you breathe much harder than normal and may include heavy lifting, digging, aerobics or fast bicycling. Think about *only* those physical activities that you did for at least 10 minutes at a time.

[Note: examples of activities may be replaced by culturally relevant examples with the same METS values – See Ainsworth et al. 2000.]

1. During the last seven days, on how many days did you do vigorous physical activities?

[Interviewer clarification: Think about only those physical activities that you do for at least 10 minutes at a time.]

[Interviewer: Include all jobs.]

_____ days

[None = 0]

[Refused = 8]

[Don't know = 9]

[Interviewer: If respondent answers 0, 8 or 9, go to Question 3.]

2. How much time in total did you usually spend on one of those days doing **vigorous** physical activities?

_____ hours ____ minutes/day

[Interviewer clarification: Think about only those physical activities that you do for at least 10 minutes at a time.]

[Interviewer probe: An average time per day is being sought. If the respondent cannot answer because the pattern of time spent varies widely from day to day, ask: "How much time in total would you spend in a **usual week** doing vigorous physical activities?"

_____ hours ____ minutes/week]

Now think about activities that take **moderate physical effort** that you did in the last seven days. Moderate physical activities make you breathe somewhat harder

than normal and may include carrying light loads, bicycling at a regular pace, or doubles tennis. Do not include walking. Again, think about *only* those physical activities that you did for at least 10 minutes at a time.

3. During the last seven days, on how many days did you do **moderate** physical activities?

[Interviewer clarification: Think about only those physical activities that you do for at least 10 minutes at a time.]

[Interviewer: Include all jobs.]

_____ days

[None = 0]

[Refused = 8]

[Don't know = 9]

[Interviewer: If respondent answers 0, 8 or 9, go to Question 5.]

4. How much time in total did you usually spend on one of those days doing **moderate** physical activities?

___ hours ___ minutes/day

[Interviewer clarification: Think about only those physical activities that you do for at least 10 minutes at a time.]

[Interviewer probe: An average time per day is being sought. If the respondent cannot answer because the pattern of time spent varies widely from day to day, or includes time spent in multiple jobs, ask: "How much time in total would you spend in a **usual week** doing moderate physical activities?

___ hours ___ minutes/week]

5. Now think about the time you spent **walking** in the last seven days. This includes walking at work and at home, walking to travel from place to place, and any other walking that you might do solely for recreation, sport, exercise or leisure. During the last seven days, on how many days did you walk for at least 10 minutes at a time?

[Interviewer clarification: Think about only the walking that you do for at least 10 minutes at a time.]

[Interviewer: Include all jobs.]

_____ days

[None = 0]

[Refused = 8]

[Don't know = 9]

[Interviewer: If respondent answers 0, 8 or 9, go to Question 7.]

6. How much time in total did you usually spend walking on one of those days?

____ hours ____ minutes/day

[Interviewer probe: An average time per day is being sought. If the respondent cannot answer because the pattern of time spent varies widely from day to day, ask: "How much time in total would you spend in a **usual week** walking?"

____ hours ____ minutes/week]

7. Now think about the time you spent **sitting** on weekdays during the past seven days. Include time spent at work, at home, while doing course work and during leisure time. This may include time spent sitting at a desk, visiting friends, travelling on a bus, reading or sitting or lying down to watch television. During the past seven days, how much time in total did you usually spend sitting on a **weekday**?

____ hours ____ minutes

[Interviewer clarification: Include time spent lying down (awake) as well as sitting]

[Interviewer probe: An average time per day is being sought. If the respondent cannot answer because the pattern of time spent varies widely from day to day, ask: "What is the total amount of time you spent sitting last Wednesday?"

____ hours ____ minutes

Chapter 7.

Development of a common instrument for use of curative medical services

Agnes de Bruin and Dirkjan Beukenhorst
Statistics Netherlands, Voorburg and Heerlen, Netherlands¹

7.1 Introduction

At present, the national use of curative medical services is typically measured through either medical registrations and/or health interview surveys (HIS). Although both these approaches have their advantages and disadvantages, HIS are considered to provide a clear added value to medical registrations, especially with respect to integrated information, information on non-users of medical services, and information on services that are not or are insufficiently covered by registrations.

In general, register data are more accurate and reliable than the information gained from HIS. Owing to lack of medical knowledge and effects of memory, respondents in a HIS are not always able to correctly report diagnoses or, for instance, the length of a hospital stay. Information about patients in psychiatric hospitals can often be collected only from registrations, since many of these patients will not be able to answer HIS questions themselves. Finally, medical registrations often have complete coverage of the use of the medical service concerned, whereas HIS are always subject to sampling and non-response errors. This factor is especially relevant when phenomena with low prevalence rates are studied or when more detailed information is required.

HIS have a number of important advantages compared with medical registrations, however. Owing to the integrated way in which a large amount of health-related information is collected, it is possible to link information on the use of different types of service by a single person. For example, questions such as “How many people use more than one type of medical service in one year?” and “How many people do not use any type of medical assistance?” can be answered through a HIS. It is also possible to study the relationship between the use of different types of medical service, as well as to relate the use of medical services to demographic and socioeconomic characteristics of patients. Likewise, integration with information on health determinants and health status is possible with a HIS. These types of integrative study are mostly not possible with medical registrations, because only a limited number of background characteristics are included and because record linkage between registrations is often difficult or impossible.

HIS also have the important advantage over medical registrations that users and non-users of health services can be compared with respect to a number of relevant characteristics, such as socioeconomic status. Information on both healthy and less healthy persons not using health services is useful for policy-makers and, for example, health insurance companies.

¹ The views expressed in this document are those of the authors and do not necessarily reflect the policies or position of Statistics Netherlands or any other organization.

Finally, HIS are a relatively cheap and quick tool to fill gaps in the information available from medical registrations. In the Netherlands, for example, there do not (yet) exist complete registrations on the use of services of general practitioners, specialists, physiotherapists and alternative practitioners (Swinkels 1996a).

In summary, HIS are a relevant tool to monitor the level of and developments in the use of health services. The objective of the EUROHIS network on “use of curative medical services” was therefore to develop a common instrument to measure this indicator in HIS as part of national health information systems. The instrument should measure use by the respondents of the most common medical services (WHO 1997) and should be:

- capable of providing data useful for health policy monitoring and evaluation;
- appropriate to be routinely included in regular national population surveys;
- aimed for use in the general population in a face-to-face interview, although options for other administration methods (telephone, self-administered) would be an asset;
- relatively short and easy to score and interpret; and
- suitable for cross-national equivalent measurement.

7.2 Methodology used to develop a common instrument for use of curative services

The “use of curative services” network followed the general EUROHIS work programme and included the following steps.

Preparatory concept exploration

At the first network meeting in Leiden (9–10 April 1999), a working definition of the indicator was formulated and a preliminary selection was made of first and second priority domains (types of medical services). Within these domains, the relevant concept elements to be measured were explored.

Review of instruments currently used and definitive concept clarification

After the first network meeting, a mail survey was conducted across the Member States in the WHO European Region (the “survey of surveys”). With respect to “use of curative medical services”, information was collected on the opinions of the countries about definitions and elements to be measured, as well as the instruments and survey methodology they had used in their latest population surveys. On the basis of these data and the conclusions from the first network meeting, the indicator network made a further demarcation of the domains and domain elements to be measured in the core instrument, and the domains or elements that could be measured as optional additions.

Development of draft common instrument

At the second network meeting in Brussels (12–13 May 2000), draft questions were formulated for the selected domains. A further information source that was useful at this stage was a Eurostat working paper on coverage of health topics by surveys in the European Union (Hupkens 1997). The following criteria were used in selecting the elements to be included, i.e. they should:

- be core instruments (with a limited number of questions);
- measure the domains and elements selected;
- use the methodological information available regarding question design;
- take into account the different health care systems in countries (i.e. be appropriate for use in all countries); and

- as far as possible (when the former criteria were first met) relate to existing questions already used in surveys

Pre-testing of draft instrument

Pre-testing was conducted in 4 European Union countries and 12 countries of central and eastern Europe and the newly independent states (CCEE/NIS) during September/October 2000. Most pre-tests were qualitative, with the number of respondents ranging between 20 and 200, using in-depth cognitive interviewing (i.e. the interviews comprised not only the survey questions but also qualitative questions that concerned the interpretation of key concepts and the way in which the respondent chose and formulated his/her answers). Following discussion of the results at the EUROHIS mid-term review in November 2000, further adjustments were made to the instrument.

Field-testing of draft instrument

Only a selection of the questions were field-tested in 2001 because of the time constraints imposed by the telephone interviews that were undertaken in the three European Union countries. Several CCEE/NIS countries and Israel also tested the instrument, mostly through face-to-face interviews or by mail. The primary aim of the field-testing was to investigate the feasibility of large-scale implementation of the instrument, but it also offered the opportunity to test some qualitative aspects of the questions in an unobtrusive way.

Final recommended common instrument

Some further adjustments were made to the instrument on the basis of the field-test results. Both a long version of the instrument and several shorter versions were finalized. A minimum set of core items was selected, and recommendations were made for additional core and optional items.

7.3 Survey of surveys

Twenty-two countries, covering 32 surveys, responded to the EUROHIS survey of surveys with items relating to use of curative medical services. Detailed inventories of the information that was collected are available on request.

7.3.1 Domains

The most common domains currently included in national HIS are hospitalization, consultations with medical doctors (general practitioners or family doctors, specialists, etc.) and dental consultations (Table 7.1). In the category “other health services”, the most common domains are physiotherapist, psychotherapist/psychologist, alternative practitioner (homeopath, acupuncturist, chiropractor/manual therapist, naturopath), mental health care centre and nurse. Although domestic help and home services not related to health care are also relatively frequent in HIS, the network group considered that these domains should not be included in the “use of curative medical services” indicator, and they are therefore not considered further here.

Table 7.1. Coverage of “use of curative medical services” domains in national HIS of 22 countries^a

Domain	No. of countries	Domain	No. of countries
Hospitalization	20	<i>Other health services</i>	
Visits to outpatient department	5	physiotherapist	7
Visits to casualty ward	3	psychotherapist/psychologist	4
		dietician	2
		speech therapist	2
Consultations with doctor	15	homeopath	6
– general practitioner	11	acupuncturist	5
– specialist	11	chiropractor, manual therapist	5
		naturopath	4
Consultations with dentist	16	mental health care centre	4
		nurse	5
		home services	6
		domestic help	4
		other categories (13 types)	21

^a Data from the EUROHIS survey of surveys.

7.3.2 Reference periods

There is large variation in the reference periods used in the instruments reported in the survey of surveys. Table 7.2 provides an overview for the most common domains included in HIS. Twelve months is the most common reference period, especially for hospitalization.

Table 7.2. Reference periods used for “use of curative medical services” domains in national HIS of 22 countries^a

Domain	Reference period	No. of countries	Domain	Reference period	No. of countries
Hospitalization	2 weeks	1	Specialist	4 weeks	1
	4 weeks	1		2 months	2
	3 months	3		6 months	1
	12 months	16		12 months	6
		3 years		1	
Medical doctor	2 weeks	4	Dentist	2 weeks	2
	4 weeks	2		2 months	2
	3 months	3		3 months	2
	12 months	6		6 months	1
		12 months		9	
General practitioner	2 weeks	3			
	4 weeks	2			
	2 months	2			
	6 months	1			
	12 months	6			

^a Data from the EUROHIS survey of surveys.

7.3.3 Diversity of instruments

From the response to the survey of surveys, it was concluded that there is a large diversity of instruments for measuring use of curative medical services in the European Region. Variation was seen in the definition of domains, in the reference periods used and in the elements that are included. There is therefore little perspective for between-country comparisons by means of ex-post harmonization of present data sets (as originally planned in the EUROHIS project). The use of curative services group therefore focused on developing a “new” instrument, based on the opinions expressed in the survey of surveys, discussions in network meetings, and the available methodological literature. The national instruments collected through the survey of surveys provided a useful reference pool for formulating these question and answer categories.

7.4 Definition of concept, domains and elements in the common instrument

7.4.1 Definition of concept

The “use of curative medical services” concept was defined as the use by the population of services provided by professionals, institutions and/or companies in the curative medical and paramedical field.

7.4.2 Selection of domains

The criteria that were used to select the domains to be covered in the common instrument were the importance of the domain in terms of volume and costs, the feasibility of measurement in a HIS, actual presence in current HIS, prospects for between-country comparisons, and the availability of national registration data.

The following domains were thus selected as first priority:

- hospitalization;
- consultations with medical doctors, particularly doctors for general medical care (e.g. general practitioners) and specialists; and
- consultations with dentists.

Although hospital discharges are registered in all countries, hospitalization was given first priority because the register data are usually admission-based rather than patient-based. HIS have an added value in measuring the number of admissions per person per year. In addition, the possibility of analysing the information in relation to socioeconomic variables and other health-related variables from HIS was considered an advantage. Disadvantages of the HIS approach for this domain are the large sample size required (because of the relatively infrequent use of services at an individual level) and the potential for bias when only non-institutionalized persons are interviewed. The latter problem is, however, predominantly restricted to the older age groups.

The domain “visits to outpatient departments for diagnostic tests” was initially suggested by the network as an optional component. After further evaluation of the pre-test results, however, it was decided not to include this because it overlapped with the questions on hospitalization and consultations with medical doctors.

As second priority domains, the following “other” health services were selected:

- physiotherapist;
- dietician;
- speech therapist;
- alternative practitioner (homeopath, acupuncturist, chiropractor/manual therapist, naturopath, other);
- nursing care at home;
- outpatient mental health care institutions; and
- psychologist or psychotherapist outside of outpatient mental health care institutions.

Although the use of dietician and speech therapist services was infrequently mentioned in the survey of surveys, they were included because they are usually part of the official paramedical health services. The instrument that was developed to measure these second priority domains was constructed in such a way that other domains can be easily added if countries deem it necessary for their national health information systems.

7.4.3 Definition of first-priority domains

Hospitals

The term “hospital” was defined to include general hospitals, single-specialty (categorical) hospitals, university (teaching) hospitals and psychiatric hospitals.

All countries that responded to the survey of surveys on this domain considered that these types of hospital should be included. Although some countries also mentioned nursing homes and institutes for the mentally disabled, it was decided to exclude these because institutionalized populations are not usually included in national HIS.

Medical doctors

It was decided that the measurement of consultations with medical doctors should specifically include consultations with doctors for general medical care and consultations with specialists. As there are differences between countries in the types of doctor that provide general medical treatment, it was decided that consultations with medical doctors should include, besides general practitioners, doctors in accident and emergency centres, doctors in occupational health care, paediatricians (in a questionnaire targeted at children) and specialists. Although in some countries gynaecologists also play a role in general medical treatment for women, it was decided not to enquire about these in a separate category. They are, however, included in an optional question on the type of specialist last consulted.

Specifying different types of doctor in the questions has the advantage that the same concept of “medical doctor” is measured, and that differences in overall frequencies can be better interpreted. Ultimately, it was decided that these specific questions should be preceded by a general “catch-all” question on consultations to all types of medical doctor (see section 7.5.2). This is also currently practised by many countries that responded to the survey of surveys (15 out of 22 have a catch-all question).

The term “medical doctors” was thus defined as including medical doctors of any kind, and to specifically include general practitioners (family doctors), doctors in accident and emergency centres, doctors in occupational health care, paediatricians and specialists. The term “specialist” refers to a doctor who has done further training in a particular specialty, either surgical (e.g. general surgeon, orthopaedic surgeon) or medical (e.g. internist, neurologist, respiratory physician, etc.).

Dentists

From the surveys of surveys, it was concluded that orthodontics is practised by both general dentists and orthodontists in most countries, and only by orthodontists in some countries. To construct a comparable instrument, it was therefore considered appropriate to include orthodontists in the definition of a dentist. Consultations with a dentist were therefore defined as consultations with (general) dentists and orthodontists.

7.4.4 Selection of elements to be measured

For the *first-priority domains*, the instrument should measure at least frequency of consultations or admissions by type of medical provider or institution. The other elements to be included differed for each domain. Table 7.3 summarizes the elements that were selected for each domain, and the initial priority given to them. Some of the core elements mentioned here were later made optional.

The selection of elements was largely based on the opinions expressed in the survey of surveys. For example, in view of the rising trend for day treatment, it was decided that day patient admissions should be included in the common instrument for hospitalization, separately from overnight admissions (inpatient care). Most countries also considered that duration of inpatient hospitalization should be measured (for day patients this is implicit in the frequency measurement). Some elements were not specifically asked for in the survey of surveys, but were considered important by the EUROHIS network, e.g. the element “presence of own teeth” for the dentist consultations. This element was included for comparability reasons: the presence of own teeth varies between countries and is an important determinant of the number of consultations with a dentist.

For the *second-priority domains* (“other health services”) it was decided to measure only whether or not the respondent had used the respective health service in the last 12 months. This was due partly to the lower priority of these domains, but also to the insufficient numbers of positive responses that would be obtained from a general population if frequencies were measured.

Besides the core elements described above, some optional elements were also suggested by the EUROHIS network (see Table 7.3). Further elements that were discussed but not included were the distinction between first and re-admissions to hospital or first/follow-up consultations with medical doctors (too difficult for the respondents), and the distinction between private and public institutions/practices (too country-specific). Other suggestions, such as referrals (to and from other health care providers) and waiting lists, were not selected for the common instrument because they were considered too specific in the context of the main objective of the common instrument. Although HIS are a unique measurement tool for the element “satisfaction with medical care”, this was excluded as it was considered to be outside the remit of the “use of curative medical services” indicator.

Table 7.3. Elements selected for domains of “use of curative medical services” indicator, and preliminary division into core and optional elements

Domain	Element	Preliminary priority
Hospitalization	Frequency and duration of inpatient hospitalization	Core
	Frequency of day patient hospitalization	Core
	Reason for hospitalization	Core
	Duration of hospitalization by reason	Optional
	Operations	Optional
Consultations with medical doctor	Frequency	Core
	Percentage of persons consulting in 12 months	Core
	Reason for consultation	Core
	Place of consultation (GP)	Core
	Type of specialist	Core
	Illness or health complaint that led to the consultation	Optional
Consultations with dentist	Frequency	Core
	Percentage of persons consulting in 12 months	Core
	Reason for consultation	Core
	Presence of own teeth	Core
	Regularity of dental consultations	Optional
Consultations with other health services	Percentage of persons consulting in 12 months	Core

7.5 Development of draft common instrument

The elements summarized in Table 7.3 were further defined and operationalized into a draft set of questions and answer categories, as described below.² The performance of the resulting draft instrument was assessed in the EUROHIS pre-test in 2000 and the larger field-test in 2001. The final recommended instrument is presented in section 7.7 and Appendix 7.1.

The instrument was designed primarily for use with adults in face-to-face or telephone interviews. In these types of (computer-assisted) survey, complex routing of questions can be easily implemented and reference periods can be further clarified by adding dates in the wording of the questions (“so since ...*(date)*...”). Such formulations cannot be used in self-administered questionnaires. Adaptations are therefore necessary when the instrument is used in postal surveys.

The instrument is suitable for answering by proxy, when necessary. This has the advantage that potential respondents who are currently hospitalized or too ill to participate in the survey can still be included, thus giving a more representative sample. When children are included in the target population, the questions should be answered by proxy and some of the questions should be adapted. Some suggestions for this are described in the instructions to the final recommended instrument (the instrument has, however, not yet been tested on a child population).

7.5.1 Draft items on hospitalization

² A detailed technical report is available from the authors on request.

Introductory text

It was felt necessary to include an introductory text in order to explain the term “hospitalization”. According to the definition that was adopted (see section 7.4.3), this covered admissions to “all” types of hospital, where general, categorical (single specialty), university and psychiatric hospitals should all be included. Since it was not appropriate to name all these different types of hospital in the questionnaire, however, it was simply mentioned that “all hospitals” should be included. According to the survey of surveys, there are only a few countries where people often use the term “clinic” when referring to hospitals; they are advised to change the introductory text to include “all hospitals and clinics”.

The introductory text initially excluded (certain types of) long-stay hospital, but this was sometimes difficult to translate and did not stop respondents from reporting admission to long-stay hospitals. This exclusion was therefore dropped; nevertheless, in countries where stays in “sanatoria” or “kurkliniken” are common phenomena, especially for non-medical reasons, it is advised to explicitly exclude these. If the instrument does result in data from long-stay hospitalization (this will be an erroneously low number anyway, as most surveys do not include institutionalized populations), the data can be adjusted accordingly. For example, cross-national comparisons can be made of the length of stay for people who report one stay within the past year, and long-stays (>30 nights) can be excluded.

It was acknowledged that that former psychiatric patients might be underrepresented in the final survey sample.

Given the high frequency of hospitalizations due to childbirth, and in view of the differences in birth rates and place of delivery (hospital or at home) between countries, it was considered necessary to explicitly include hospitalizations due to childbirth in the introduction. To allow separate analysis of this later, hospitalization due to childbirth was also specifically mentioned in the items on reason for hospitalization.

Inpatient care and day patient care

The administrative rules that differentiate day patient care from inpatient care differ from country to country, thereby making it impossible to construct a common survey instrument that will meet the needs of all countries. While many countries use midnight (0.00 hours) as the defining point, others use, for example, noon (12.00 hours). For the common instrument, it was decided to choose the most frequently used time, midnight (i.e. inpatient admission is equivalent to having an overnight stay), and to clearly define this in the question wording so that the same concept would be measured in different countries. To distinguish day patient care from outpatient care, it was decided to include in the definition of day patient care that the patient should have been admitted to a hospital bed. On the basis of the pre-test results, the word “admitted” was further emphasized and the term “for treatment” was deleted to prevent confusion with outpatient diagnostics.

Inpatient care was thus defined as a stay in hospital overnight or longer, and day patient care as admission to a hospital bed with discharge before midnight the same day.

It may be a problem for some countries that, while the national administrative system may not regard hospital admission for childbirth as an inpatient admission, the survey data may code it as such (e.g. where hospitalization for childbirth includes midnight it will be counted as an inpatient admission, even if length of stay is under 24 hours). If these countries find it essential to deduce the numbers of childbirth hospitalizations, then a follow-up question (devised according to the national definition) could be inserted after the question on reason for inpatient admission.

Frequency of hospitalization, reference period

The preferred reference period for hospitalization is 12 months because of the low prevalence rate of hospitalization in a typical survey sample. Underreporting due to memory effects is probably minor, as people do not tend to forget hospital admissions (Hupkens 1997, Swinkels 1996b). Furthermore, most countries already use a reference period of 12 months and their experiences are mostly good, e.g. in comparisons with register data.

To solve the problem of hospital stays that cross the boundaries of the reference period (i.e. stays that started more than 12 months ago but ended less than 12 months ago, and stays that have not yet ended at the time of interview), it was decided that only hospital stays that *ended* during the reference period of 12 months would be counted. In the case of proxy interviewing, it has to be noted that present hospitalizations should be excluded. Furthermore, by using the wording “Have you been in hospital ...” instead of “Have you been admitted to hospital ...”, those hospital stays that started more than 12 months ago and ended within the reference period are implicitly included.

Frequency of hospitalization was thus defined as the number of inpatient stays and the number of day patient admissions that ended in the past 12 months.

Duration of inpatient hospitalization (volume)

In determining length of stay in hospital, it is generally advised to count the number of nights, as this is less ambiguous than the number of days (Hupkens 1997). It is then possible to count the total number of nights (of all hospitalizations in the past 12 months), the number of nights in each separate hospitalization, or only the number of nights related to the last hospital stay. It is generally preferred (Hupkens 1997) to count the number of nights for each hospitalization separately by asking for admission and discharge dates, but this was considered to be inappropriate for a common instrument. Asking only for the duration of the last hospitalization has the disadvantage that it underestimates hospitalization of frequent service users. Furthermore, the total number of nights spent in hospital during the last year is then less accurately derived at the individual level (by multiplying the number of nights of the last stay by the number of hospitalizations in the past year). The latter has the consequence that correlation studies with other variables will underestimate the “real” effects.

It was therefore decided to measure the duration of inpatient hospitalization by asking for the total number of nights of all hospital stays that ended in the past 12 months.

At the aggregate level, this measures the volume of hospitalization (number of nights) in 12 months. The nights of current hospitalizations are not counted, but this is compensated for by including all the nights of stays that started more than 12 months ago (and ended in the past 12 months). The possible error that is introduced by asking the respondent to add up all the nights spent in hospital during several stays, without specifying dates, is accepted. However, as the large majority of people hospitalized within the last year have only had one stay, few respondents will need to count the number of nights of several stays.

The instrument can also be used to estimate the average stay of a hospitalization by dividing the total number of nights of the ended hospitalizations by the number of ended hospitalizations in the past 12 months.

If countries wish to measure the duration of hospitalization by the reason for hospitalization, they could add an extra question for this (note that this would be an optional question and not part of the core instrument).

Reason for hospitalization

Some of the countries that responded to the survey of surveys considered that the reason for hospitalization (divided into broad categories) should be included in the common instrument. However, such information must be kept very general owing to the small number of respondents who have been hospitalized; furthermore, it may be difficult for respondents to place their diagnosis in the most appropriate category.

It was therefore decided that only the categories “childbirth” and “injury or accident” should be asked for in specific items, while other reasons would be included in a follow-up question (later classified as an optional item) with an open-answer category. If countries want to investigate the reason for hospitalization more fully, they will need to code the answers afterwards, preferably by trained health personnel and using either the International Classification of Diseases (ICD) (WHO 1992) or the International Classification of Primary Health Care (ICPC) (Lamberts & Wood 1998). An electronic ICPC version is available that is specific for use with electronic patient records and for research purposes, and that can be converted to the ICD10 (Okkes et al. 2000).

For practical reasons, it was decided that the reason for hospitalization should be asked only for the last inpatient stay that ended during the past 12 months and for the last day patient admission in the past 12 months.

To allow a rough estimate of the frequency of surgical intervention, a separate question on this issue could be added to the last (ended) hospitalization. This is an optional addition, not belonging to the core instrument.

7.5.2 Draft items on consultations with medical doctors

Introductory text

It is essential to clarify which contacts with doctors are to be included and which are not. For the common instrument, “consultations with medical doctors” were defined so as to include: (a) only consultations on behalf of the respondent him/herself (proxies are allowed for children); (b) consultations at the doctor’s practice, during house-calls (for a general practitioner) and by telephone; (c) consultation with a doctor in an outpatient department, and (d) consultation with a doctor in a casualty ward or emergency centre. They should exclude: (a) contacts with the sole purpose of making an appointment and (b) consultations during hospitalization. The result is a rather full introductory text, but this was considered a necessary evil.

Medical doctors

The preliminary draft questionnaire included only visits to the specific types of doctor that were selected by the EUROHIS network (see section 7.4.3); an overall frequency of consultation could then be derived by summing these individual frequencies. However, the pre-test revealed great divergence between countries in the definitions for these types of doctor that would likely result in incomparable data. The solution was sought in a general “catch-all” question on consultations with medical doctors of any type, which was then followed by the more specific questions. The list of doctor types may not be exhaustive for all countries, but the first “catch-all” question should ensure data on all types. Furthermore, a respondent-friendly routing (filter) is possible, whereby respondents who have had no consultations in the past 12 months can skip all the questions on the specific doctors, while those who have consulted a doctor in the last 12 months but not the last four weeks can skip all the “four-week” questions on the specific doctors.

The field-test revealed that respondents had problems in differentiating between the general question on (all) medical doctors and the following question on general

practitioners. This was solved by inserting a short introductory sentence prior to these questions. For translation purposes and instruction of interviewers, the definition of “specialist” was further specified (see section 7.4). For practical reasons, it was decided that the type of specialist seen would only be asked of the last consultation within the past 12 months. The pre-coded answer categories for the type of specialist are based on common categories in the EUROHIS survey of surveys. Consultations with other specialists can be obtained from the open answer category.

Reference periods

The relatively long reference period of 12 months, which is commonly used in survey questions on medical consultations (see Table 7.2), was considered appropriate only for measuring whether or not a consultation had taken place. In this context, it can be used to calculate the percentage of persons consulting in the last 12 months (participation rate), as well as to investigate the reason for the last consultation (in the last 12 months).

To calculate frequency of consultations, however, a much shorter reference period is needed because of memory effects. A reference period of two weeks has been recommended for general practitioner consultations, to minimize the effect of underreporting forgotten consultations and overreporting consultations outside the reference period (telescoping). However, in view of the fact that the common instrument should also measure consultations with other, less frequently consulted types of doctor, a two-week reference period would yield insufficient numbers considering the average survey sample size. Furthermore, it was felt necessary to use the same reference period for the different types of doctor so as to avoid respondent confusion. For these reasons, the network decided, after lengthy discussion, to use a reference period of four weeks.

Reason for consultation

As in section 7.5.1 on hospitalization, it was decided to enquire about the reason for medical consultation only in relation to the last consultation in the past 12 months. Only the main reason is asked and the answers are pre-coded, the categories being derived from examples in the EUROHIS survey of surveys. For practical reasons, the same answer categories are used for each type of doctor, although some categories may be less relevant for some types of doctor in some countries.

The answer categories start with “accident or injury”, followed by a general category on other health problems (“illness or health complaint”). The next category, “check-up”, was initially followed by a category “preventive examination or test”, but it became apparent in the pre-test that the distinction between this and “check-up” was not clear, so it was dropped. The “check-up” category was further clarified by adding that it was unconnected with illness or complaints. Inclusion of the “administrative procedure” category is particularly relevant for international comparisons, as this is an important reason for consultations (e.g. in connection with sick leave) in some countries.

An optional open question could be added to obtain information on the underlying illness or health complaint that led to the consultation, with the use of the ICD or ICPC coding systems.

7.5.3 Draft items on consultations with dentists

Introductory text

In the pre-test it became clear that explanatory text was needed for the first question on presence of own teeth. The text addition “answer ‘yes’ if you have one or more teeth left” was deleted after the pre-test, but was shown to be necessary in the field-test. It was therefore moved into the interviewer instructions.

Reference periods

For reasons of consistency and comparability, it was decided to use the same reference periods as selected for medical consultations, i.e. four weeks for measuring frequency and 12 months for the percentage of persons consulting a dentist/orthodontist and for reason for last consultation.

Reason for consultation

Pre-coded answer categories were derived from the EUROHIS survey of surveys. “Orthodontic regulation (brace)” was included as a category because of the specific nature of this treatment and because consultations with orthodontists are included in the definition of dentist consultations. Based on the pre-test results, the category “due to pain” was replaced by “pain or complaints”, and a category “for denture” was added.

Regularity of dental consultations

An optional question could be added on the regularity of consultations with dentists. This question is specifically directed to the work of general dentists and does not refer to orthodontists. If countries find it relevant for their national information needs, they could also add an optional item on whether respondents are invited by their dentist for routine check-ups.

7.5.4 Draft items on use of other health services

Specification of health services

For some of the health services it was difficult to choose the question wording, as the organization and nomenclature of these services often differ between countries. It was therefore decided to specify the health services as far as possible in the common instrument, such as in the case of alternative practitioners. Countries should use equivalent national wordings wherever relevant.

The common instrument consists of a core list of health services. Other health services may be added to the list according to national needs, but these should be placed *after* the last item of the core list.

Reference period

As the percentage of persons consulting the health services is the only element to be measured, a reference period of 12 months was chosen, as for medical and dental consultations.

7.6 EUROHIS field-testing in 2001

The instrument on use of curative medical services had to be reduced in size for the purposes of the field-test (where the whole survey instrument should not take longer than 30 minutes). Only the following topics were therefore field-tested: frequency of inpatient

and day patient hospitalizations; length of stay; and frequency of consultations with medical doctors and dentists. Certain domains (e.g. consultations with specific types of doctor) gave more problems than others (e.g. hospitalization), although not every country had problems. Many problems were reported in France, while none were reported in Germany.

7.6.1 Hospitalizations

The questions on inpatient and day patient hospitalization posed few problems in the field, although it is still not certain whether respondents correctly distinguished between inpatient and day patient care. If countries wish to investigate this issue further, they are advised to ask not only the minimum core items but also some extra questions, e.g. on reason for hospitalization (as recommended in Appendix 7.1) and possibly on the duration of each inpatient hospital stay. “Incorrectly” reported hospital stays (e.g. long stays) can then be removed during data cleaning.

7.6.2 Consultations with medical doctors and dentists

Many differences exist between countries in types of medical doctor and the reasons for consulting them. This makes comparison of the total volume of such consultations difficult. The draft instrument therefore includes a “catch-all” question, which refers to consultations with doctors in general, followed by a series of questions that ask about consultations with specific types of doctor. In the three European Union field-testing countries, the “catch-all” question on the total number of consultations in the past four weeks was not used as a filter for proceeding to the specific questions (i.e. even if no consultations in four weeks were reported, respondents were still asked the specific four-week questions). This offered an opportunity to study memory effects, by comparing the frequencies of the “catch-all” question with the summed frequencies of the specific questions. The appropriateness of using a filter question could also be assessed on the basis of these data. In the CCEE/NIS field-tests, the “catch-all” question was used as a filter, as recommended in the draft instrument.

It was expected that the use of the “catch-all” question would produce a lower prevalence than if the specific questions on different subtypes were used. Table 7.4 supports this, and shows that consultation frequency in the previous four weeks based on the “catch-all” question was underestimated in the European Union countries, e.g. by 4% in Germany and about 10% in the United Kingdom (unweighted data). In the CCEE/NIS countries, of which the Czech Republic is included in Table 7.4 as an example, there were also lower prevalence rates using the “catch-all” question, but the differences were smaller here because of the use of a filter. The network group considered that the level of underestimation that was produced by the “catch-all” question was acceptable, and was moreover counterbalanced by the gains in cross-population comparability. Analysis of the data of the “non-filtered” European Union data further showed that, in total, only two respondents reported no consultations on the catch-all question, and at the same time reported one or more on the specific questions. It is therefore recommended that the “catch-all” question indeed be used as a filter in order to shorten the average interview duration.

Table 7.4 also shows the advantage of collecting cross-national data on consultations with specific types of doctor. It suggests that in the Czech Republic, probably in common with other CCEE, doctors at the workplace are far more frequently consulted than in the European Union countries, although the overall frequency of consultation is

about the same as that in Germany. In the United Kingdom, specialists appear to be less frequently consulted (and general practitioners more frequently) than in the other countries. The doctor-specific questions (and those referring to reason for consultation) thus provide important information for interpreting differences in overall volume.

Table 7.4. Number of medical doctor consultations (in last four weeks) from the “catch-all” question compared with the number of consultations from the questions on specific types of medical doctor

Country	Unweighted						Weighted					
	All doctors	Specific questions					All doctors	Specific doctor types				
		GP	AE	WP	Spec	Total		GP	AE	WP	Spec	Total
United Kingdom	193	164	8	3	38	213	193	158	14	3	29	204
Germany	494	258	4	3	245	510	494	274	4	3	245	526
France	415	275	9	9	159	452	410	292	9	8	157	466
Total	1102					1175	1097					1196
Czech Republic	494	267	15	24	194	500						

GP: general practitioner; AE: accident and emergency doctor; WP: doctor at workplace; Spec: specialist. Net sample size = 400 in each country.

If respondents in the field-test had not consulted a medical doctor within the previous four weeks, they were asked whether the last consultation had been within the last 12 months or longer ago than 12 months, or had never occurred. The aim of this question was to allow the computation of an annual consultation rate (percentage of persons with one or more consultations in 12 months), as well as to generate more respondents for possible follow-up questions on reason for consultation. The field-test interviewers were asked to observe respondent behaviour and make a note as to whether respondents (*a*) asked for clarification or interrupted the interviewer, and (*b*) had doubts about their answer or had problems with choosing an answer category.

In France and the United Kingdom, respondents quite often interrupted the interviewer and/or asked for clarification, while the questions appeared to function well in Germany. The problems in France appeared to be related to translation errors, but the data suggested that the United Kingdom respondents were more hesitant about their answers. The difficulty in remembering the date of last contact is probably explained by the relatively low consultation rates in the United Kingdom, suggesting that the questions that refer to the “last four weeks” may provide more reliable data than those referring to the “last 12 months”. However, recall problems did not appear to be a major problem in the 12-month questions, and it is therefore recommended that these questions be maintained in the common instrument, with the advantage that the same time period is used for all medical services.

7.7 Recommended common instrument for use of curative medical services

The final recommended common instrument is a “long version” which includes a minimum set of 10 core items, additional core items recommended for inclusion, and a number of optional items (See Table 7.5 and Appendix 7.1). Countries may, of course, add further

elements to the common instrument according to their own national priorities. For comparability reasons, such items should be inserted *after* the common items.

Table 7.5. Overview of core and optional items in the final instrument, indicating those included in the EUROHIS pre-test and field-test (item numbers refer to Appendix 7.1)

Domain	Minimum core items	Additional core items	Optional items	Pre-test	Field-test
<i>Inpatient hospitalization</i>					
Frequency in 12 months	1, 2			X	X
Total duration in 12 months	3			X	X
Duration of last stay			4e	X	
Reason for last stay		4a, 4b		X	
Other reasons for last stay (open question)			4c	X	
Operation during last stay			4d		
<i>Day patient hospitalization</i>					
Frequency in 12 months	5, 6			X	X
Reason for last stay		7a, 7b		X	
Other reasons for last stay (open question) ^a			7c	X	
Operation during last stay ^a			7d		
<i>Consultations with medical doctors (in general)</i>					
Frequency in four weeks	8				X
Consultation in past 12 months	9 (and 8)				X
<i>Consultations with general practitioner</i>					
Frequency in four weeks		10		X	X
Consultation in past 12 months		11 (and 10)		X	X
Reason for last consultation ^b		12a		X	
Place of last consultation			12b	X	
<i>Consultations with doctor in accident and emergency centre/casualty</i>					
Frequency in four weeks		13		X	X
Consultation in past 12 months		14 (and 13)		X	X
Reason for last consultation ^b		15		X	
<i>Consultations with doctor at workplace</i>					
Frequency in four weeks		16		X	X
Consultation in past 12 months		17 (and 16)		X	X
Reason for last consultation ^b		18		X	
<i>Consultations with medical or surgical specialist</i>					
Frequency in four weeks		19		X	X
Consultation in past 12 months		20 (and 19)		X	X
Reason for last consultation ^b		21a		X	
Type of specialist last consulted			21b	X	
<i>Consultations with dentist</i>					
Frequency in four weeks	23			X	X
Consultation in past 12 months	24 (and 23)			X	X
Reason for last consultation		25		X	
Presence of own teeth	22			X	X
Regularity of consultations			26	X	
<i>Use of other health services</i>					
Use in past 12 months			27	X	

^a Allows assessment of respondents' ability to distinguish between day patient hospitalization and outpatient visits.

^b With pre-coded answer categories. An open question, "What was the underlying illness or health complaint?" can be added after this item.

When analysing the data from the common instrument, it is recommended that the following approaches be used to derive various frequency and consultation rates.

7.7.1 Inpatient hospitalization

Percentage of persons with inpatient hospitalization in one year: [Number of respondents with inpatient hospitalization in past 12 months (item 1) – Number of respondents with one stay that were still hospitalized (“0 stays” in item 2)] / Total number of respondents (item 1).

Number of inpatient hospitalizations in one year per person in the population: Total number of inpatient stays that ended in past 12 months (item 2) / Total number of respondents (item 1).

Number of inpatient hospitalizations in one year per hospitalized person: Total number of inpatient stays that ended in past 12 months (item 2) / Total number of respondents with one or more stays (item 2).

Number of nights in hospital in one year per person in the population: Total number of nights of all hospital stays that ended in the past 12 months (item 3) / Total number of respondents (item 1).

Number of nights in hospital in one year per hospitalized person: Total number of nights of all hospital stays that ended in the past 12 months (item 3) / Total number of respondents with one or more stays (item 2).

Average duration of an inpatient hospitalization: Total number of nights of all hospital stays that ended in the past 12 months (item 3) / Total number of inpatient stays that ended in the past 12 months (item 2).

7.7.2 Day patient hospitalization

Percentage of persons with day patient hospitalization in one year: derived directly from item 5.

Number of day patient hospitalizations in one year per person in the population: Total number of day patient stays (item 6) / Total number of respondents (item 5).

Number of day patient hospitalizations in one year per hospitalized person: Total number of day patient stays (item 6) / Total number of respondents with one or more stays (item 6).

7.7.3 Consultations with doctors and dentists

The following measures can be derived for items 8/9, 10/11, 13/14, 16/17, 19/20 and 23/24, where each pair consists of a four-week item followed by a one-year item.

Percentage of persons consulting in four weeks: derived directly from the four-week item. This measure provides the most valid estimation of doctor/dentist consultation rates. However, the next measure (annual consultation rate) is useful when presenting consumption rates for different medical services.

Percentage of persons consulting in one year: (Number of respondents with consultation in past four weeks + Number of respondents without consultation in past four weeks but with last consultation less than 12 months ago) / Total number of respondents (four-week item).

Number of consultations in one year per person in the population: (Total number of consultations in past four weeks × 13) / Total number of respondents (four-week item).

Number of consultations in one year per person with consultation: (Total number of consultations in past four weeks × 13) / (Number of respondents with consultation in past four weeks + Number of respondents without consultation in past four weeks but with last consultation less than 12 months ago).

7.7.4 Use of other health services

Percentage of persons using other health services in one year: derived directly from item 27.

7.8 Conclusions

The development of a common instrument for use of curative medical services has been a difficult endeavour, given the differences in the structures of the European health systems and in the way in which they are used. Furthermore, the diversity of existing survey instruments made it necessary to develop a new instrument for which no prior methodological evidence was available. Despite these obstacles, the EUROHIS network group has searched for common denominators and comparable definitions and has developed an instrument that can be used in all countries. In seeking to balance the need for a simple instrument that would nevertheless produce relevant and comparable data, not all problems could be completely solved but “best solutions” were formulated. The resulting instrument performed reasonably well in the field-test. The instrument is constructed in such a way that, in addition to the 10 core items, countries can select further core items and/or optional items, and can also add country-specific elements when this is desired. It is expected that the instrument will yield data that will allow international comparisons of the use of health care services and will also assist in the interpretation of actual differences.

References

- HUPKENS, C. (1997) *Coverage of health topics by surveys in the European Union*. Luxembourg, EUROSTAT (Working Paper Population and Social Conditions, 3/1998/E/No. 10).
- LAMBERTS, H. & WOOD M. (1998) *International Classification of Primary Care*, 2nd ed. Oxford, Oxford University Press.
- OKKES, I. ET AL. (2000) ICPC-2-E: the electronic version of ICPC-2. Differences from the printed version and the consequences. *Family practice*, **17**: 101–107.
- SWINKELS, H. (1996a) *Measuring medical consumption: registrations or health interview surveys? Paper presented at the PHARE Seminar on Health Interview Surveys, 28 May – 6 June 1996*. Voorburg, Statistics Netherlands.

- SWINKELS, H. (1996b) *Health interview surveys and the measurement of medical consumption. Paper for PHARE Seminar on Health Interview Surveys, 28 May – 6 June 1996.* Voorburg, Statistics Netherlands.
- WORLD HEALTH ORGANIZATION (1992) *International Statistical Classification of Diseases and Related Health Problems*, 10th revision. Geneva, World Health Organization.
- WORLD HEALTH ORGANIZATION (1997) *Common methods and instruments for health interview surveys in Europe. Report on the 4th WHO Consultation, 26–28 February 1997.* Copenhagen, WHO Regional Office for Europe (document EUR/ICP/INFO 02 03 05).

Appendix 7.1
EUROHIS recommended common instrument for use of
curative medical services
(finalized after field-testing in 2001; note that questions 4, 7, 12,
15, 18, 21 and 25–27 were not included in the field-test
questionnaire)

Instructions for survey managers

Technical details

Introduction before item 1: in countries where the term “clinic” is often used by people in referring to hospitals, it is advisable to mention that “All hospitals and clinics are included”. In countries where stays in a sanatorium or “kurklinik” for non-medical reasons often occur, these stays may be specifically excluded in the introductory text.

The questionnaire is designed for face-to-face and telephone interviews (CAPI and CATI). For self-administration (mail surveys), the layout should be adapted.

The instrument is also suitable to be answered by proxy, although it would need adapting, i.e. replacing the “you” format in the items by the person’s name; item 4 should also be changed to “Thinking of the last ended stay in hospital as an inpatient, ...”.

Items 5–7 can be omitted in countries where day hospitalization does not exist, while 7b can be omitted in countries where hospitalization for childbirth always includes a night.

An optional item regarding the underlying illness or health complaint of (the last) doctor consultations can be added after items 12a, 15, 18 and 21a: “What was the underlying illness or health complaint?” (with open answer category). This question should be omitted when the answer on the first item (12a, 15, 18, 21a) is “check-up”.

Notes for translation

The wording in the instrument that is presented here is final, as recommended after pre-testing and field-testing earlier versions.

Whenever relevant, countries should use national equivalent wordings of the health care providers or services mentioned in items 21b and 27.

Item 13: the term “accident or emergency centre” should only be translated if such centres exist separately from hospital casualty departments.

Item 16: “doctor at your workplace” refers to doctors in occupational health services.

Item 23: “on your own behalf” means that the respondent him/herself was examined or treated.

Adaptations for children (0–14 years)

If children are included in the target population, the items should be answered by proxy (see “technical details” above). In the introductory text before item 1, and in items 4b and 7b, the term “for giving birth/to give birth” should be replaced by “for birth”. In items 16–18, the term “doctor at your workplace” should be replaced by “paediatrician”. In items 19–21, the term “medical or surgical specialist” should be replaced by “medical or surgical specialist other than a paediatrician”.

Making shorter versions of the questionnaire

It is possible to include only parts of the instrument in a survey, depending on the national priorities, although it is recommended that the minimum set of core items always be included (see Table 7.5). It is advisable to maintain the same order of items, however, and additional country-specific items should always be inserted *after* the respective core items.

If the doctor-specific questions (items 10–21) are not included, an additional item on the reason for the last doctor consultation may be inserted after item 9 (only to be answered if 8 > 0, and 9 is “less than 12 months ago”), i.e. “Thinking of the last time you consulted a doctor, what was the main reason for the consultation?” Answer categories as in item 12a.

QUESTIONNAIRE
(long version, 15+ years)

Interviewer: The next questions are about hospitalization. All hospitals are included. [for females up to age 50 years, add:] Hospitalizations for giving birth should also be included.

1. During the past 12 months, that is since (*date one year ago*), have you been in hospital as an inpatient, that is overnight or longer?

Yes

No → 5

2. How many separate stays in hospital as an inpatient have you had since (*date one year ago*)?
Count all the stays that ended in this period.

..... stays [if still in hospital = 0 stays → 5]

3. How many nights in total did you spend in hospital during this (these) inpatient stay(s)?

..... nights

4. Thinking of the last time you stayed in hospital as an inpatient:
a. Were you in hospital because of an injury or accident?

Yes → 4d

No → 4b if respondent is a woman aged 15–50 years;
otherwise → 4c

- b. Were you in hospital to give birth?

Yes → 4d

No

- c. For what illness or complaint were you hospitalized?

.....

d. Did you undergo an operation?

Yes

No

Question e only if more than one hospital stay in question 2:

e. How many nights were you in hospital during your last stay?

..... nights

5. During the past 12 months, that is since (*date one year ago*), have you been admitted to hospital as a day patient, that is admitted to a hospital bed, but not required to remain overnight?

Yes

No → 8

6. How many times have you been admitted as a day patient since (*date one year ago*)?

..... times [= number of days; count each day]

7. Thinking of the last time you were in hospital as a day patient:

a. Were you in hospital because of an injury or accident?

Yes → 7d

No → 7b if respondent is a woman aged 15–50 years;
otherwise → 7c

b. Were you in hospital to give birth?

Yes → 7d

No

c. For what illness or complaint were you hospitalized?

.....

d. Did you undergo an operation?

Yes

No

Interviewer: The next questions concern consultations with medical doctors. Include visits at a doctor's practice, house-calls as well as consultations by telephone. Also count visits to doctors at outpatient or emergency departments, but exclude contacts during hospitalization.

(Interviewer instruction: contacts with the sole purpose of making an appointment should not be counted)

Interviewer: The next question is about the use of other health services.

27. During the past 12 months, that is since (*date one year ago*), have you visited or used any of the following health care providers or services for your own health needs?

Physiotherapist	Yes/No
Dietician	Yes/No
Speech therapist	Yes/No
Homeopath	Yes/No
Acupuncturist	Yes/No
Chiropractor, manual therapist	Yes/No
Naturopath	Yes/No
Other alternative practitioner	Yes/No
Nursing care at home	Yes/No
Institute for outpatient mental health care	Yes/No
Psychologist or psychotherapist, outside outpatient mental health care institutes	Yes/No

Chapter 8.

Development of a common instrument for use of medicines

Timo Klaukka and Jaana Martikainen

Research and Development Centre, Social Insurance Institution (KELA), Helsinki, Finland

8.1 Introduction

The use of medicines is increasing in Europe along with the growing number of elderly people, the increasing health expectations of the population and medical advances. A major problem that faces various countries is the continual increase in drug expenditure, caused mainly by new drugs replacing older, less expensive ones. Medicines are often valuable tools in the control of diseases, but their overuse causes extra costs and adverse effects. On the other hand, underuse still occurs owing to limited financial resources.

Under these circumstances, the rational use of medicines is an essential objective. To evaluate the quality of drug therapy, information is needed on the extent and the characteristics of drug use in various population groups. Health interview surveys (HIS) are an important contribution to the information system on drug utilization. They are good sources of information to complement registers and sales statistics. The methodology used in different countries should be identical, however, to improve the comparability of the data.

The EUROHIS use of medicines network has developed a new series of questions on the use of medicines, based on the information that was gathered from the “survey of surveys”. This survey identified 20 national health surveys from 17 countries that included questions about use of medicines. The number of questions on medicines ranged from 1 to 14, but many of the questions contained several items. Although the approaches in the 20 studies varied somewhat, many strategies were found that were useful in the development of a new instrument for use of medicines.

Preliminary tests of the new draft instrument were performed in several European countries. After the modification of the questions, more extensive field-testing was performed in three European Union countries: France, Germany and the United Kingdom. The final instrument is now ready for others to use.

8.2 Why study the use of medicines?

An instrument that measures the use of medicines is relevant for several reasons.

- It is important to know the proportion of the population that is exposed to various groups of medicines. This exposure should be measured with respect to age, gender and socioeconomic group, and the studies should be repeated at regular intervals to investigate trends in the use of medicines.

- Information on the role of medicines in the treatment of major diseases should be available. The use of certain medicines may reveal previously unknown patterns of morbidity.
- The extent of self-medication is a relevant study topic, as self-medication may substitute for the use of prescription medicines. The same applies to the use of herbal medicines (on which WHO conducts a periodic survey), which may be an alternative or a supplement to established treatment, and may also reflect the attitude of the population towards the official health services system.
- Vitamin supplements and minerals are important to study, as these preparations are increasingly used for maintaining and enhancing good health (neither of which is their original or official purpose). The present use indicates health promotion behaviour, and it may correlate with other forms of health-related lifestyle.

8.3 What is a medicine?

To study drug utilization, it is first necessary to specify what is meant by a “medicine”. Although the European Union has a legal definition, it appears to be interpreted differently among WHO Member States, and the classification of the various substances into categories differs from one country to another. While substances that need a sales licence are often officially categorized as medicines, traditional drugs of herbal origin do not usually need a sales licence and thus lack the necessary documentation for registration. In contrast, homeopathic substances are now officially defined as medicines by European Union legislation. Furthermore, lay people often perceive all these types of substance as medicines because they are believed to relieve symptoms and to improve poor health.

Health interview surveys collect data from lay people and thus provide a useful basis for defining medicines. A review of the questions used in national surveys revealed, however, that there may be no need to explain to the respondent what is meant by a “medicine”. Several surveys simply enquired whether the respondent had been using “any medicines” during a certain period of time, while some surveys gave examples, such as tablets, drops, suppositories, injections, eye-drops or plasters. There were, however, certain product categories that in some studies were mentioned separately, probably because they otherwise tend to be (rightly or wrongly) either included or excluded from the answers. Examples of such categories are vitamins, minerals, nutritional supplements, natural/herbal remedies, contraceptive pills and estrogen preparations. The inclusion of these in (or their exclusion from) the responses can greatly affect the comparability of the resulting data.

After reviewing the concepts that were currently used in health surveys, the working group defined a medicine as follows:

Medicine is a product that is used to alleviate symptoms, to prevent illness, or to improve poor health, and which is ordinarily purchased from a pharmacy.

8.4 How can the use of medicines be measured?

It was clear from the survey of surveys that different countries use different approaches to measure the use of medicines in the population. The most frequently used method is to provide a list of specified diseases (or complaints) and/or purposes of the use. The respondent then checks off the categories as appropriate. Another approach is to ask for the brand name of the medicine that the respondent has been using.

8.4.1 List of specified items

Medicine use can be measured by asking the respondent to respond to a list of diseases or symptoms (e.g. medicines for asthma, cholesterol-lowering drugs) and/or by inquiring about the use of predefined drug groups (e.g. sleeping pills, analgesics). Such a list usually includes common diseases or medicine groups with a relatively high rate of use. The sample size used will greatly influence the possibility of identifying users of medicines in these groups.

The survey of surveys revealed 44 different items that were mentioned as the purpose of the medication or as a group of medicines. The items included in at least five different surveys were as follows (in order of frequency):

- pain relievers (12), often specified for rheumatism, musculoskeletal disorders, etc.
- anti-hypertensives (10)
- sedatives, tranquillizers (10)
- heart (cardiovascular) medicines (8), in some studies further specified, e.g. heart failure, angina pectoris, etc.
- laxatives (7)
- anti-diabetic medicines (7); in one study insulin and anti-diabetic tablets were listed separately
- sleeping pills (7)
- penicillin or other antibiotics (6); also just antibiotics
- anti-allergy medicines (6); also antihistamines
- vitamins (6); also together with minerals or as restorants
- anti-asthmatics (5)
- antidepressants (5), also called “stimulants”
- cholesterol-lowering agents (5)
- medicines for digestive disorders (5).

In the context of the EUROHIS study, it was noted that there could be some overlap between questions on use of medicines and questions on preventive services, as the latter may also enquire into aspects of medicine use, e.g. vitamins, cholesterol-lowering drugs and anti-hypertensives. If it is intended to collect both sets of information in a survey, it would therefore be wise to check for significant overlap.

8.4.2 Recording medicines by brand name

In five surveys, the respondents were asked to report the brand names of the medicines they were using at the time of the study or had used during a defined period of time. This approach enables the medicines to be categorized according to an official classification, such as by ATC group. On the other hand, the same medicine can be used for various purposes (e.g. aspirin for pain or to prevent thromboembolism; beta blocking agents for hypertension or for migraine). Thus the reporting of the product's name does not always reveal the purpose of the use by the respondent.

Information on the names of medicines is more reliable when a self-completed questionnaire is used or when an interview is conducted at the respondent's home, where the containers or prescription forms are easily available.

8.4.3 Recall period

Taking a tablet is not a particularly special event, and thus the recall period used in a survey on use of medicines must be relatively short. On the other hand, the longer the period, the greater the sample size for further analyses.

The most frequently occurring (8 surveys) recall period in the survey of surveys was 14 days (two weeks). Other alternatives were “at present”, “during two days”, “during four weeks”, “during six months” and even “ever”. The EUROHIS network group considered that the two-week approach would be most appropriate in order to reduce memory problems. It was accepted that this recall period was shorter than that used in the EUROHIS instrument on use of curative medical services.

8.5 Development of a draft instrument

The working group could not find any existing instrument that could be recommended for use across Europe, although many countries had useful approaches to measuring the use of medicines. The group therefore decided to construct a new instrument on the basis of the questions already in use in European countries and then to test this instrument further. The following questions were constructed for field-testing.

1. Have you taken any prescribed medicines (including contraceptive pills or other hormones) during the last two weeks?

Yes

No [GO TO 4]

2. (If yes) Were they medicines for:
 - A. high blood pressure?
 - B. other cardiovascular disease?
 - C. lowering the blood cholesterol level?
 - D. diabetes?
 - E. pain in the joints (arthrosis, arthritis)?
 - F. headache or migraine?
 - G. other pain?
 - H. asthma?
 - I. chronic bronchitis or emphysema?
 - J. allergic symptoms (eczema, rhinitis)?
 - K. depression?
 - L. stomach troubles?

Or were they:

M. sleeping tablets?

N. antibiotics such as penicillin?

O. (for women in fertile age – assumed 50 years or younger) contraceptive pills?

P. (for women in or after menopausal age – assumed 45 years and older) hormones for menopause or osteoporosis?

Q. some other medicines prescribed by a doctor? (If yes): What type of medicines? _____

3. (OPTIONAL) What were the names of the prescription medicines that you took during the last two weeks?

Medicine 1 ... Medicine 10.

4. Have you taken any medicines (including vitamins and minerals) not prescribed by a doctor during the last two weeks?

Yes / No

5. (If yes):Were they medicines for:

A. pain?

B. cold, flu or sore throat?

C. allergic symptoms (eczema, rhinitis)?

D. stomach trouble?

Or were they:

E. vitamins, minerals or tonics?

F. some other medicines not prescribed by a doctor? (If yes): What type of medicines? _____

6. (OPTIONAL) What were the names of the medicines not prescribed by a doctor that you took during the last two weeks?

Medicine 1 ... Medicine 10.

8.6 EUROHIS pre-testing 2000

The draft instrument on use of medicines was tested together with the draft instruments on quality of life, chronic physical conditions, use of preventive health care and use of curative medical services in 15 European countries (Bosnia and Herzegovina, Bulgaria, Czech Republic, Estonia, Finland, Hungary, Italy, Latvia, Lithuania, the Netherlands, Romania, the Russian Federation, Slovakia, Spain and Ukraine) in August–October 2000. The instrument was also tested alone in a small sample in Germany.

No problems were apparent in the translation of the use of medicines items. The draft instrument worked relatively smoothly in the pre-test and there were few comments on the use of medicines section. Nevertheless, it was pointed out that the division between prescription and non-prescription drugs can differ between countries, and thus the results would not be directly comparable. In addition, doctors may prescribe medicines that have a non-prescription status. One proposed solution to this was that the first set of questions on use of prescription medicines should be interpreted as referring to all medicines prescribed by a doctor or used on a doctor's initiative or recommendation. The second set of questions about non-prescription medicines would then refer to the use of medicines independently of a doctor. It was agreed that these explanations should be included in the instructions to the interviewers.

At the EUROHIS mid-term review in November 2000, the draft instrument was presented to managers of national health surveys from 30 European countries. Some of the managers expressed concern about the regional variations in comprehending the concept of a medicine. Furthermore, elderly respondents may not always know the conditions for

which they are taking medicines. As a solution to the latter concern, it was suggested that information be collected at the level of brand names.

8.7 EUROHIS field-testing 2001

The use of medicines instrument was field-tested in France, Germany and the United Kingdom in nearly original form. Information on the names of the medicines used was not collected, however, because there were major difficulties in determining a common data base for coding these names. Different countries can use different brand names for the same preparations, and thus a central data base would be difficult to establish. ATC classification is not used in all European countries, and it is sometimes laborious to find a category for medicines that are available only in one or two countries.

There were no major technical problems in the field-testing of the instruments, and the results were mainly as expected. Use of medicines appeared to be very common among middle-aged respondents and especially the elderly. A majority of the European Union respondents (aged 35–75 years) had taken prescribed medicines in the two weeks before the interview, i.e. 52% of men in France and the United Kingdom and 61% of men in Germany. The respective figures for women were 68%, 59% and 71%. The somewhat higher proportions in Germany were mainly due to the older age structure of the German sample compared to those in the other two countries.

The finding that women used prescribed medicines more frequently than men has been shown previously in several studies and statistics. The proportion of users increased with age in all three countries, but not completely systematically. This slight inconsistency might have been caused by the relatively small numbers of respondents. Education did not show a systematic correlation with the level of use of prescribed medicines.

The medication pattern was similar to the pattern seen in sales statistics. The most frequently used medicines in all three countries were anti-hypertensives and various types of analgesics. The female dominance in the frequency of use of prescribed medicines did not apply for all types of drugs. More men than women used anti-hypertensives, other cardiovascular medicines, cholesterol-lowering drugs and anti-diabetics, while the reverse was true for analgesics and various psychotropic drugs. The use of oral contraceptives and menopausal hormones was very common in the relevant age groups, and this explained in part the generally higher level of use of medicines among women.

The validity of the answers was crudely evaluated by comparing the results from the use of medicines instrument with those of the chronic conditions instrument, as some of the groups of medicines that were enquired about were related to the chronic diseases that were listed. The major difference between these two instruments was the recall period, which was two weeks in the use of medicines instrument but 12 months in the chronic conditions instrument. The respondents in all three European Union countries gave similar answers to the corresponding questions in the two instruments. Ideally, all the respondents who answered “Yes” to the use of specific medicines within the previous two weeks should also have responded positively to having the corresponding chronic condition in the previous 12 months. This kind of validity check showed 100% correspondence for some of the items and somewhat lower correspondence – but still high, around 90% – for others.

The use of non-prescription medicines was not as frequent as that of prescription drugs. The proportion of respondents who had taken medicines not prescribed by a doctor during the previous two weeks ranged from 20% (French men) to 45% (British women). Of the various subgroups of over-the-counter medicines, vitamins, minerals and tonics were by far the most frequently reported in all three countries. Other frequently taken medicines were analgesics and treatment for the common cold. There were no internal controls in the

questionnaire that would allow the results of the use of over-the-counter medicines to be checked. In general, non-prescription medicines are used temporarily, while many chronic conditions require long-term medication. The use of a relatively short recall period meant that respondents who took regular medication were more likely to be identified than respondents who took medicines on a short-term basis. The clear difference in the proportion of users of prescription versus non-prescription medicines therefore appeared to be logical.

The comments from the field-testing indicated that respondents were hesitant in answering some of the questions, but these appeared to be minor problems and there was no apparent need to alter the choice of items or the question wording.

8.8 Conclusions

The use of medicines instrument performed well with respect to translation of the items and in pre-testing and more extensive field-testing. The number of questions is realistic and the instrument would not demand too much space in a survey. Some critical comments on the methodology were made during the process, but better solutions were not identified. A major methodological deviation from the original idea was the exclusion of collecting the names of the medicines used. In national surveys, however, this might still be a good idea if an up-to-date classification of medicines is available and if the survey team itself is able to code and interpret the results. It is recommended that the instrument presented in Appendix 8.1 be used as a set of core questions to improve the basis for international comparisons of the use of medicines.

Appendix 8.1

EUROHIS recommended common instrument for use of medicines (finalized after field-testing in 2001)

1. Have you taken any prescribed medicines (including contraceptive pills or other hormones) during the last two weeks? *[Note: this question refers to all medicines that have been prescribed by a doctor or used on a doctor's initiative or recommendation.]*
 - Yes
 - No [GO TO 3]

2. (If yes) Were they medicines for (Yes/No for each item):
 - A. high blood pressure?
 - B. other cardiovascular disease?
 - C. lowering the blood cholesterol level?
 - D. diabetes?
 - E. pain in the joints (arthrosis, arthritis)?
 - F. headache or migraine?
 - G. other pain?
 - H. asthma?
 - I. chronic bronchitis or emphysema?
 - J. allergic symptoms (eczema, rhinitis)?
 - K. depression?
 - L. stomach troubles?

Or were they:

 - M. sleeping tablets?
 - N. antibiotics such as penicillin?
 - O. (for women in fertile age – assumed 50 years or younger) contraceptive pills?
 - P. (for women in or after menopausal age – assumed 45 years and older) hormones for menopause or osteoporosis?
 - Q. some other medicines prescribed by a doctor? (If yes): What type of medicines? _____

3. Have you taken any medicines (including vitamins and minerals) not prescribed by a doctor during the last two weeks? *[Note: this question refers to all medicines that have been used independently of a doctor.]*
 - Yes / No

4. (If yes): Were they medicines for (Yes/No for each item):
 - A. pain?
 - B. cold, flu or sore throat?
 - C. allergic symptoms (eczema, rhinitis)?
 - D. stomach trouble?

Or were they:

 - E. vitamins, minerals or tonics?
 - F. some other medicines not prescribed by a doctor? (If yes): What type of medicines? _____

Chapter 9.

Development of a common instrument for use of preventive health care

Marijke W. de Kleijn-de Vrankrijker and Wim Davidse
TNO Institute of Preventive Health Care, Leiden, Netherlands

9.1 Introduction

Although we know that not all disease and illness can be prevented, health promotion and the use of preventive health services can reduce morbidity and mortality, and consequently also reduce health care expenditure and the social effects of ill health. The availability of comparable cross-national data is an important element in the development of policies for preventive health at national and international levels.

In the context of the EUROHIS project, “use of preventive care” was interpreted in a broad and non-traditional way, taking into account risk factors, disease, injury, congenital conditions, and disability according to the International Classification of Functioning, Disability and Health (WHO 2001), as well as interventions provided by health professionals and actions taken by individuals.

In developing the instrument on use of preventive health care, the following phases of work were undertaken:

- concept exploration and choice of main domains
- selection of topics within domains
- review of instruments already used in European countries
- further elaboration of topics
- EUROHIS pre-test 2000
- EUROHIS field-test 2001
- production of a recommended instrument for use of preventive health care.

9.2 Concept exploration and choice of main domains

The first network meeting was held on 10 April 1999 in Leiden, the Netherlands. The participants agreed that, while the classical definition of prevention (primary, secondary and tertiary) is often used in health care, for the purpose of the EUROHIS project it would be more useful to make working definitions in terms of the conditions/situations (disease, disability, accidents, etc.) to be prevented.

This led to a discussion as to whether or not complications of diseases should be included. For example, while everyone agreed that preventing the complications of hypertension is a preventive measure, most participants considered that prevention of pressure sores in a chronically ill person is an aspect of care. On this basis, it was decided that prevention of complications of diseases should not be covered by the common instrument.

In addition, although all participants agreed that cost–effectiveness and prevalence are important aspects of preventive health services, it was decided that they were outside the remit of the common instrument and thus would not be incorporated into the working definition of use of preventive services.

In the context of the EUROHIS project, it was decided that preventive care should include activities/interventions that focus on:

- prevention and control of risk factors;
- prevention and early detection of disease;
- prevention of injuries and congenital conditions; and
- prevention of impairments, disabilities and handicaps in terms of the International Classification of Impairments, Disabilities and Handicaps (WHO 1993); according to the International Classification of Functioning, Disability and Health (WHO 2001) this is equivalent to the prevention of disability (i.e. impairments, activity limitations and participation restrictions).

Furthermore, the use of preventive health care would be defined as that provided by professionals, as well as the use of specific devices and measures used or taken by individuals.

The main domains to be included would be:

- immunization and vaccination programmes
- health promotion
- medication/dietary supplements
- screening
- specific programmes to improve quality of life in individuals with existing health problems.

9.3 Selection of topics within domains

The aspects to be measured within each domain were considered on the basis of the following criteria:

- importance of the health problem to public health and prevention;
- prevalence of the health problem;
- possibility of developing questions for use in regular national population surveys;
- new health problems and challenges relevant for prevention; and
- topics not covered by other EUROHIS indicators.

The topics selected within each domain were as follows.

Immunization/vaccination

- Influenza (elderly and other risk groups)
- Diphtheria (total population); poliomyelitis, rubella and tuberculosis (children)

Health promotion

- Nutrition habits and change of weight (thus omitting physical activity, smoking, alcohol, use of safety belts and hypertension)

Medication/supplements

- Hormone replacement (thus omitting use of calcium supplements and vitamin D)

Screening

- Breast cancer
- Cervical cancer
- Hypertension
- Hypercholesterolaemia
(thus omitting dental check-ups and prenatal screening).

At a later EUROHIS Board meeting in Rome on 29 May 1999, it was decided to consider two further aspects:

- Rehabilitation as a structural programme
- Prevention of occupational health problems in general.

9.4 Review of instruments already used in European countries

In 1999/2000, WHO conducted a postal survey (the “survey of surveys”) in order to make an inventory of the instruments currently used (or under development) in the European Region to monitor the EUROHIS indicators. Concerning the use of preventive health care, the following specific questions were asked.

- Which of the selected preventive programmes (including “rehabilitation as a structural programme” and “prevention of occupational health problems in general”) exist in your country?
- Which health professionals (e.g. general practitioners or family doctors, specialists in hospitals, preventive or public health specialists, nurses, other) or institutions (e.g. hospitals, preventive health services, other) are involved?
- Who are the target groups (general population, risk groups, specific age groups, etc.)?
- What information is available in your country about these preventive programmes (number of participants and participation rate, information about non-response, evaluation)?

Replies relating more generally to preventive health care (i.e. organization, use of preventive programmes, etc.) were received from 12 countries: Austria, Belgium, the Czech Republic, Denmark, Germany, Hungary, Italy, Lithuania, the Netherlands, Norway, Spain and Ukraine. The results of the survey were discussed at the second meeting of the indicator network, held in Brussels in May 2000.

The most important conclusions were as follows.

- Concrete questions referring to many of the selected topics do exist in a number of national health interview surveys.
- Some preventive programmes exist in almost all countries, e.g. screening for cervical and breast cancer and immunizations in general. In most countries these programmes are population-based.
- Some programmes, such as screening for high cholesterol, rehabilitation programmes and prevention of occupational health problems, have not yet been implemented in most countries, and estrogen replacement as a preventive measure is used in only one or two countries.

- In most countries, curative professionals (general practitioners, specialists and hospital nurses) carry out preventive activities. In a number of countries, public health specialists are involved in some programmes; usually they work in hospitals but sometimes in specialized public health services.
- From the information received it was concluded that rehabilitation programmes should not be considered part of preventive health care.
- Few countries have public health monitoring of the coverage/participation in preventive programmes.
- Target groups for preventive programmes are similar in different countries.

Some common patterns could be identified with respect to target groups.

- Screening for breast cancer is generally carried out in women of 40–50 years and up to 70 years of age, although in some countries the target group is much younger.
- Screening for cervical cancer is generally carried out in women between 25–30 and 55–60 years of age.
- Screening for hypertension is provided in some countries for people aged 40 and over, while in others it is provided for those over the age of 50 or 60.
- Screening for hypercholesterolaemia is generally carried out in those aged 35–40 years and over.
- Influenza vaccination is usually offered to those at risk and to the elderly (60–65 years and over).
- Immunization of children is carried out between 0–9 and 0–12 years of age.

9.5 Further elaboration of topics

Actual questionnaires that contained items on prevention were received from 19 countries: Austria, Belgium, Denmark, Finland, Germany, Hungary, Ireland, Israel, Italy, Latvia, Lithuania, the Netherlands, the Russian Federation, Slovakia, Slovenia, Spain, Sweden, Switzerland and the United Kingdom. Most of these questionnaires had been used in national health surveys.

In designing items to enquire about the topics to be covered within each of the main domains, it was decided to use the same format as much as possible for all topics, that is:

- ever/never participated
- when participated last time
- reason (why or why not)
- on whose initiative.

9.5.1 Immunization: influenza

Immunization of the elderly (60–65 years and older) and other risk groups is very common in most countries. Indication of risk group could be derived from the chronic conditions indicator, e.g. those with diabetes, heart attack and chronic heart disease (heart failure), chronic obstructive lung disease (including bronchitis and emphysema), asthma and renal insufficiency.

In general, the general practitioner is involved in immunization, but self-vaccination is also possible. A common instrument could enquire about when immunization took place, on whose initiative this was done, and the reason for no immunization. Optional questions could enquire about knowledge and opinion concerning influenza immunization. Many of

the national health interview survey questionnaires contain items on influenza immunization.

9.5.2 Immunization: diphtheria, poliomyelitis, rubella and tuberculosis

Immunization against a number of communicable diseases is an important preventive programme in (almost) all countries and is compulsory in a few countries. Only a small number of countries have monitoring systems. The provision of public health information about immunization against diphtheria, poliomyelitis and rubella remains important. Existing questionnaires contain only a few questions about this subject. A common instrument could include questions on time of last immunization, and on whose initiative this was done. Questions about tuberculosis could be asked in countries where a tuberculosis vaccination programme for children exists.

9.5.3 Dietary habits and weight change

Many health interview surveys contain a large number of questions about this subject. The vast majority of questions yield epidemiological data that can form a basis for the development of preventive programmes. However, the data themselves cannot be considered as preventive programmes or preventive measures. The network group focused on items dealing with changes in dietary habits, the reasons for these and the role of health professionals. Some existing questions were found that were considered appropriate for inclusion in the EUROHIS instrument. It was noted that if a health examination is also carried out, then such dietary information could be linked to data on blood pressure, body mass index and cholesterol level.

9.5.4 Hormone replacement

Overuse of estrogens to treat menopausal complaints is a problem in many countries, as the risk of (especially) breast cancer is increased. Important aspects of this topic are when hormone replacement began and ended, the reasons for its use and the person who took the initiative to start it. An optional question on the brand of medication used could be included. No relevant questions were found in existing health interview surveys.

9.5.5 Screening: breast cancer

Almost all countries have some kind of screening programme (mainly X-rays; in a few cases only breast palpation, the effectiveness of which has not been proven). In most countries this is organized as a population-based programme. The target group in most countries is women aged 40–50 years and older, although in some countries the target group is much younger. The network group concluded that the questions should be asked of women aged 20 years and older, and should relate to time of last mammography, reason for mammography and frequency of subsequent check-ups. For countries with a national X-ray screening programme, additional questions on acceptance or otherwise of the invitation were recommended. Countries without such a programme could include questions about self-examination.

9.5.6 Screening: cervical cancer

Almost all countries have some kind of screening programme, mostly organized as a population-based programme. The target group is generally women aged 25–30 to 55–60 years. The network group concluded that the questions should be asked of women aged 20 years and older, and should relate to the time of last smear test, the reason for having it done and the frequency of subsequent check-ups. For countries with a national screening programme, additional questions on acceptance or otherwise of the invitation were recommended.

9.5.7 Screening: hypertension

Few countries have a national preventive programme. In general, family doctors (can) offer screening (case finding) to patients who are considered at risk, but few countries have any systematic evaluation of this. Questions recommended for inclusion in a common instrument relate to when the blood pressure was last checked, who checked it, whether or not respondents have been informed that they have high blood pressure, and what the resulting medical advice was. Other questions could relate to why the blood pressure was measured and who took the initiative to do so, as well as screening programmes at the workplace.

9.5.8 Screening: hypercholesterolaemia

Few countries have a national preventive programme. Population-based screening is not generally accepted, as it does not appear to be cost-effective. In general, family doctors (can) offer screening (case finding) to patients who are considered at risk, but few countries have any systematic evaluation of this. Questions recommended for inclusion in a common instrument relate to when the cholesterol level was last checked, whether or not respondents have been informed that their cholesterol level was too high, and what the resulting medical advice was. Other questions could ask about who took the initiative to measure cholesterol, screening programmes at the workplace, and awareness of cholesterol level.

9.5.9 Rehabilitation programmes and prevention of occupational problems

Although these are important topics, the network group did not recommend their inclusion in a common instrument. First, there are great differences across Europe in both the organization of health care systems and the meaning of the word “rehabilitation”. Second, the topics were considered too specific for a general population-based health interview survey.

9.5.10 Production of a draft common instrument

After the topics for inclusion had been chosen, a review was made of the various existing questions that dealt with each topic. Prototype questions for the draft EUROHIS instrument were then formulated. For 14 questions, the original wording was used. Nineteen other questions were based on existing questions, but did not have the same wording. This was

done in order to take account of the different health care systems in the various countries. In total, 12 questions were new.

9.6 Pre-testing 2000

After the second network meeting in Brussels in 2000, an examination was made of possible overlap between the preliminary draft instrument on use of preventive health care and some of the other draft EUROHIS instruments – for chronic physical conditions, use of curative medical services and use of medicines. Only minor problems occurred, however, and these did not lead to any changes to the draft instrument on use of preventive health care. In preparation for pre-testing, some questions were changed slightly (e.g. spelling and grammar, making response items exclusive). Pre-testing was carried out in 17 countries in August–October 2000.

The main conclusions from the pre-testing, discussed at the EUROHIS mid-term review in November 2000, were that:

- the proposed list of questions was too long;
- most respondents had difficulty in remembering childhood vaccinations and could not answer these questions;
- the wording of some questions could be improved; and
- translation into the various languages caused few problems; where necessary, equivalent translations could be found.

It was therefore suggested that the number of core questions could be reduced. A favoured option was to include only questions on “Ever/never?” and “When was the last time?” as core questions, leaving all other questions optional. In addition, questions on vaccination could be removed, retaining only the questions on influenza immunization.

As a result of the pre-testing results, the draft instrument was revised as follows:

- the number of core questions was reduced and the number of optional questions increased;
- the questions on vaccination against diphtheria, poliomyelitis, tuberculosis and rubella were removed;
- some further answer categories were added, e.g. inclusion of “olive oil” and “to live more healthily” in the change of dietary habits questions, and inclusion of “don’t know” where appropriate;
- wording was altered for some questions;
- a new question was introduced as to whether the respondent had been advised to change dietary habits;
- a further routing instruction was added; and
- where there were several answer categories, the appropriate guideline was given (i.e. “only one answer” or “more than one answer possible”)

9.7 Field-testing 2001

The core items of the draft instrument for use of preventive health care (see Appendix 9.1) were field-tested during 2001, alongside the other EUROHIS instruments. The data were analysed both quantitatively and qualitatively.

The quantitative analysis included a consistency check with items in other EUROHIS indicators (e.g. correlating the response to frequency of blood pressure measurement with the chronic conditions item that enquires about hypertension), checking that only women had answered the questions on hormone replacement and breast/cervical cancer screening, and checking the distribution of answer categories. It was concluded that the only problem was that a small number of respondents found it difficult to recall specific times of events, such as the month in which they were vaccinated against influenza, when they last had blood pressure or cholesterol measured, or when they last had hormone replacement therapy, mammography or a cervical smear test. This was a minor problem and adjustments to the questions were considered unnecessary.

Qualitative analysis revealed no problems from the interviewers' perspective. Interruptions by respondents occurred in 1–30% of cases, and most frequently for the questions asking for the length of time since the last blood pressure or cholesterol measurement, and whether hormone replacement therapy had ever been used. Other reactions (where, for example, respondents requested clarification, repetition, etc. or expressed doubt, gave an estimate or changed their answer) were seen in 1–15% of cases. Again, adjustments to the questions were considered unnecessary.

9.8 Conclusions

The work that has gone into the development of the EUROHIS indicator has been undertaken in the hope that it will contribute to cross-national comparability of data on preventive health care in Europe. For those countries that have not previously included questions on preventive health care in their national surveys, the instrument presented in Appendix 9.1 offers a way of collecting national data that will also allow international comparisons. For those countries that already conduct surveys on a regular basis, and therefore do not wish to change their questions in order to maintain comparability of data over time, the recommended instrument can act as a reference instrument. Depending on the degree of similarity between the national instrument and the recommended common instrument, data comparisons may be possible by mapping similar questions.

In both cases, the user should keep in mind that the instrument as such has not yet been fully validated.

It should be noted that the recommended instrument consists of both core and optional items. For practical reasons, the optional questions were not included in the field-test but they are included in this report for the convenience of the potential user.

References

- WORLD HEALTH ORGANIZATION (1993) *International Classification of Impairments, Disabilities and Handicaps (ICIDH)*. Geneva, World Health Organization.
- WORLD HEALTH ORGANIZATION (2001) *International Classification of Functioning, Disability and Health (ICF)*. Geneva, World Health Organization.

Appendix 9.1
EUROHIS recommended common instrument for use of
preventive health care
(finalized after field-testing in 2001; note that the optional
questions were not included in the field-test questionnaire)

Interviewer: Now I would like to ask you some questions about influenza vaccination

1. Have you ever been vaccinated against influenza?

Yes

No → Q3

Don't know → Q3

2a. When were you last vaccinated against influenza? Was it (*code the first that applies*)

in 2001?

in 2000?

in 1999 or before? → Q3

Don't know → Q3

2b. In what month?

.....Month (1...12. Don't know = 99)

Optional questions

I. Were you last vaccinated (*only one answer*):

of your own accord,

on the advice of a doctor or other health professional,

on the advice of the employer or workplace physician,

on the advice of the local public health service, or

on the advice of someone else?

II. Why have you not been vaccinated against influenza? Is it (*more than one answer possible*):

because you didn't know how to go about it?

because you didn't think about it?

because you thought it wasn't worth while?

because you thought vaccination can be dangerous?

because you didn't know one can be vaccinated?

for some other reason?

Interviewer: Now I would like to ask you some questions about eating habits

3. Have you changed your eating habits in the past 3 years?

Yes

No → Q5

Don't know → Q5

4. What have you changed in your eating habits?

a. Do you now eat butter and margarine rather more, about the same or rather less than before?

b. Do you now use olive oil rather more, about the same or rather less than before?

c. Do you now eat fat in general rather more, about the same or rather less than before?

d. Do you now eat meat and sausages rather more, about the same or rather less than before?

e. Do you now eat fish rather more, about the same or rather less than before?

f. Do you now eat fruit, vegetables and salad rather more, about the same or rather less than before?

g. Do you now eat wholemeal products like muesli or wholemeal bread rather more, about the same or rather less than before?

h. Do you now eat pasta rather more, about the same or rather less than before?

i. Do you now use salt rather more, about the same or rather less than before?

j. Do you now use sugar rather more, about the same or rather less than before?

k. Do you now eat eggs rather more, about the same or rather less than before?

l. And in general do you now eat rather more, about the same or rather less than before?

(Answer categories: "rather more", "about the same", "rather less" and "don't know")

Optional questions

III. What was the main reason for changing your eating habits? Was it (*only one answer*):

to lose weight,

to maintain weight,

to live more healthily,

because of a disease or health problem, or

for some other reason?

IV. Has anybody advised you to change your eating habits?

Yes

No → Q5

Don't know → Q5

V. Who advised you to do that? Was it (*more than one answer possible*):

a doctor, nurse or other health professional,

a dietician,

someone else, or

of your own accord?

Interviewer: Now I would like to ask you some questions regarding blood pressure

5. Have you ever had your blood pressure measured?
 Yes
 No → Q7
 Don't know → Q7
6. When was the last time you had your blood pressure checked? Was it
 less than 3 months ago,
 3–5 months ago,
 6 months to 1 year ago,
 1–3 years ago, or
 more than 3 years ago?
 Don't know

Optional questions

- VI. Who measured your blood pressure last time?
 Was it a doctor or nurse or other health professional,
 did you do it, or
 did another layperson measure it?
- VII. When checking your blood pressure, have you ever been told that you had high
 blood pressure?
 Yes
 No → Q7
 Don't know → Q7
- VIII. Because of your high blood pressure, did a doctor advise you to (*more than one
 answer possible*):
 reduce salt intake?
 lose weight?
 decrease alcohol intake?
 increase physical activity?

Interviewer: The next questions are about cholesterol

7. Have you ever been checked for blood cholesterol?
 Yes
 No women → Q9; men → finish
 Don't know women → Q9; men → finish
8. When did you last have your blood cholesterol measured? Was it:
 less than 3 months ago,
 3–5 months ago,
 6 months to 1 year ago,
 1–3 years ago, or
 more than 3 years ago?
 Don't know

Optional questions

IX. When measuring your blood cholesterol, has anyone ever told you that it was too high?

Yes

No

women → Q9; men → finish

Don't know

women → Q9; men → finish

X. Because your cholesterol was too high, did a doctor advise you then to (*more than one answer possible*)

take drugs?

lose weight?

change nutrition?

increase physical activity?

(Questions 9–16 are only for women 20 years of age and older; otherwise →finish)

Interviewer: The next questions are about hormone replacement

9. Have you ever used a hormone (estrogen) replacement therapy (tablets, injections, gels or transdermal patches), excluding for reasons of contraception?

Yes

No

→ Q13

Don't know

→ Q13

10. In what year did you start?

..... year (1900...2001. Don't know = 99)

11. Are you still using it?

Yes → optional questions XI & XII

No

12. In what year did you stop using it?

..... year (1900...2001. Don't know = 99)

Optional questions

XI. Why did (do) you use it? (*more than one answer possible*)

for the treatment of complaints?

for the prevention of heart disease?

for the prevention of osteoporosis?

for the prevention of other complaints?

for some other reason?

- XII. Did you start using it (*only one answer*):
 of your own accord,
 on the advice of your doctor, or
 on the advice of a nurse or another health professional?

Interviewer: The next questions are about screening for breast cancer

13. Have you ever had a mammography, that is an X-ray of one or both of the breasts?

Yes
 No → Q15
 Don't know → Q15

14. In what year did you have your last mammography (breast X-ray)?
year (1900...2001. Don't know = 99)

Optional questions

- XIII. What was the reason for this last mammography (breast X-ray)? (*more than one answer possible*)

Did you notice something not quite right in your breast yourself (e.g. a lump)?
 Did your GP notice something not quite right in your breast during an examination?
 Did your GP advise it, without there being something wrong?
 Did your gynaecologist advise it without there being something wrong?
 Because of breast cancer in your family?
 By invitation from a national or local screening programme?
 Any other reason?

- XIV. At what age did you have your first mammography (breast X-ray) done?
 years. (Don't know = 99)

- XV. How often did you have check-ups after the first mammography (breast X-ray)? (*only one answer*):
 at least once a year,
 every two years,
 every three years,
 about once in five years,
 less often than once in five years, or
 never?

(Questions XVI to XVIII are for countries with a national breast X-ray screening programme and women over 20 years; otherwise → Q15)

- XVI. Have you been invited to a screening for breast cancer during the past two years?
 Yes
 No → Q15

XVII. Did you go to have an examination because of this invitation?

Yes → Q15

No

XVIII. Why not? Was it (*more than one answer possible*):

because you thought it wasn't necessary?

because you find such examinations unpleasant?

because you did not have time for it?

because you have already been treated or had already an operation?

because you already have regular check-ups?

because you had recently had an X-ray?

because you are shortly going to have an X-ray?

for some other reason?

Interviewer: Now I would like to ask you some questions about cervical smear tests

15. Have you ever had a cervical smear test?

Yes

No → finish

Don't know → finish

16. In what year was your last cervical smear taken?

..... year (1900...2001. Don't know = 99)

Optional questions

XIX. What was the reason for this last cervical smear? Was it (*only one answer*):

because of symptoms,

because you visited a gynaecologist,

by invitation from a national or local programme,

for another medical reason, or

for another reason (not especially medical)?

XX. At what age was your first cervical smear taken?

..... years (Don't know = 99)

XXI. How often did you have check-ups done after the first cervical smear? (*only one answer*)

at least once a year,

every two years,

every three years,

about once in five years,

less often than once in five years, or

never?

(Questions XXII to XXIV are for countries with a national cervical cancer screening programme and women of 20 years or older; otherwise → finish)

XXII. Have you been invited to a screening for cervical cancer during the past five years?

Yes

No → finish

XXIII. Did you go to have an examination because of this invitation?

Yes → finish

No

XXIV. Why not? Was it (*more than one answer possible*):

because you thought it wasn't necessary?

because you find such examinations unpleasant?

because you did not have time for it?

because you have already been treated or had already an operation?

because you already have regular check-ups?

because you had recently had a cervical smear?

because you are shortly going to have a cervical smear?

for some other reason?

Chapter 10.

Development of a common instrument for quality of life

Mick Power

University Hospital of Edinburgh, United Kingdom

10.1 Introduction

The phrase “quality of life” is used in many different ways, and one of the major issues that faces this area of work is how the term should be defined and conceptualized. One of the key distinctions that has been made is that between health-related and non-health-related quality of life (e.g. Spilker 1996). The starting point for a number of the health-related definitions has been the well known World Health Organization (1948) definition of health as “a state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity”.

The inclusion of wellbeing in the WHO definition has led some researchers to focus too narrowly on self-reported psychological wellbeing as being the only aspect of quality of life of importance (e.g. Dupuy 1984). However, “wellbeing” has to be seen as the narrower term; it is an important aspect of quality of life (e.g. Bech 1996), but is not the only aspect that needs to be considered. The challenge has been to specify the range of health-related and non-health-related aspects of quality of life that should also be included, such that “quality of life” is not simply another term for “wellbeing”.

The WHO definition of health has clearly provided an excellent starting point for defining quality of life (e.g. WHOQOL Group 1995), but it leaves open two key questions. First, what other areas should be included in addition to the physical, mental and social ones? And second, should the conceptualization include, for example, objective characteristics of the individual in addition to the individual’s subjective evaluation? The existing definitions and measures take many different approaches to these two questions, as will be illustrated subsequently when three of the most widely used measures are described. Nevertheless, there may now be an emerging consensus for both of these key issues. In addition to the physical, mental and social aspects there is now a recognition that spiritual and religious aspects need to be included in health-related quality of life (e.g. Power et al. 1999; Spilker 1996; WHOQOL Group 1995) and a range of aspects of the individual’s physical environment needs to be included in non-health-related quality of life.

The term “quality of life” is now widely used in the health indicator and health outcome literature. There are a considerable number of definitions of the term, but the definition that has been proposed by the World Health Organization (WHOQOL Group 1995) is:

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 is a broad-ranging concept affected in a complex way by the persons’ physical health, psychological state, level of independence, social

relationships and their relationship to salient features of their environment.

Given the increasing use of the concept in all areas related to health, there is a clear need for a scale that can be used to compare subgroups within any particular culture, but which could also be used to make comparisons between different cultures.

10.2 Methodology

The development of the quality of life (QOL) indicator for the EUROHIS project consisted of five main steps:

- review of existing QOL instruments – the “survey of surveys”
- choice of QOL instrument
- development of the draft common instrument – items and response scales
- translation of items
- EUROHIS pre-testing 2000
- EUROHIS field-testing 2001.

The analyses summarized in this report have been carried out on the existing data sets that were collected with one or more versions of the WHO measure of QOL, the WHOQOL. The main purpose of the analyses was to explore the possibility that a small number of WHOQOL items might be used to characterize QOL in a manner that could be used in national health interview surveys. A preliminary summary of the findings from the EUROHIS field trial is also included.

10.3 The survey of surveys

The preliminary review of instruments used in the WHO European Region provided data on 25 national health interview surveys (HIS). The replies showed that there were a wide variety and range of QOL questions that were currently in use in national HIS, although a proportion of these were based on existing published instruments. Eleven of the surveys were found to use QOL instruments or questions developed for their own specific purposes. Four surveys reported that they did not measure QOL at all. The SF-36 was reported as being used in seven surveys, the WHOQOL in three surveys, and the EuroQol in three surveys.

The results thus reflect the conceptual diversity in the area. SF-36 is used more widely than WHOQOL and EuroQol, although the numbers of surveys are too small to estimate how significant this difference is. This is also of secondary importance, as the three instruments measure different, albeit very close, concepts. This was confirmed by the finding that several surveys use a number of instruments alongside each other. For example, the Danish HIS uses items related to the WHOQOL, the SF-36 and the EuroQol. The Belgian HIS applies both SF-36 and a questionnaire on social wellbeing, while the Netherlands HIS plans to use a combination of its own questionnaires and the SF-12. This seems to support the view that the related but different concepts under the umbrella “quality of life”, and the different purposes and uses of QOL data, are not likely to be completely covered by a single instrument, but rather by a combination of instruments.

According to the survey respondents, the most important requirements for a QOL instrument for use in a HIS were “easy to understand” and “easy to score and interpret”. A

little less important was that it is “short” and “used in as many countries as possible”. Usefulness for resource allocation was considered less relevant.

10.4 Choice of QOL instrument

The existence of established instruments with widely differing conceptualizations of QOL led to the decision that one possible way to devise a set of questions would be to use a subset of questions from one of the existing measures. The fact that three measures provided the basis for QOL measurement in over half of the national surveys suggested that the choice of instrument could be limited to a consideration of these three.¹

The three instruments under consideration, the SF-36, the EuroQol and the WHOQOL, each take contrasting approaches to the conceptualization of QOL. In November 1999, therefore, a meeting was held in Barcelona at which representatives for the three different measures made presentations. Only the main arguments will be summarized very briefly here; a fuller report of the meeting is available (WHO 2000).

- The SF-36 is acknowledged to be the most widely used measure of QOL and has the advantage of the existence of a large number of different language versions. Weaknesses of the measure include the fact that it was developed in one culture, that there are so-called “floor” and “ceiling” effects with the range of the measure, and that it mixes together both objective and subjective items. However, in terms of its use in the EUROHIS programme, the main problem was that the group was informed that the measure would be substantially revised in the near future. If the EUROHIS project were therefore to choose the SF-36 as the basis of its QOL indicator, there was a danger that the items chosen would be quickly superseded by the new instrument.
- The EuroQol is another widely used measure that is available in a considerable number of European languages. Within the EuroQol, QOL is defined as the absence of problems across a number of items and domains. This conceptualization of QOL contrasts with the approach taken by the SF-36 and by the WHOQOL. In the words of the original WHO definition “health is not merely the absence of ...”, so the reviewing group considered that, although the EuroQol might be a useful instrument for measuring change in clinical trials, it would not suffice for general population surveys. In psychometric terms, the positive end of the scale is collapsed into a zero or “absence” value within the EuroQol, which may produce considerable ceiling effects in normal healthy populations.
- The WHOQOL was the most recently developed of the measures under consideration. The WHOQOL takes a consistently subjective approach to the estimation of QOL, it considers both positive and negative aspects of items, and has the major advantage of having been developed in several different cultures simultaneously. An additional practical advantage with the WHOQOL was the possibility of access to a substantial data set with data collected from many different European cultures. Although these data were collected from opportunistic rather than representative samples, the possibility of immediate access to these data sets would allow some of the steps in the development of the indicator to be omitted.

On the basis of the information that had been presented to the group at the Barcelona meeting, and following subsequent deliberations by the EUROHIS indicator network

¹ The WHO-Five wellbeing scale (Bech et al., in press) was also discussed by the QOL working group, but was considered less useful in the current context owing to its unidimensional nature.

group, it was decided to take the WHOQOL as the starting point for the selection of items for the EUROHIS indicator on QOL.

10.5 Development of draft common instrument – items and response scales

The choice of a particular instrument, the WHOQOL, provided the EUROHIS group with the possibility of a different method of item and response scale selection than for many of the other indicators. That is, the existence of a range of data sets from European and non-European respondents allowed the choice of a statistical and conceptual method of selecting putative items for inclusion in the indicator, rather than having to generate the items and response scales afresh as in the case of some of the other indicators. Data collected with the 26-item short form of the WHOQOL, the so-called WHOQOL-BREF, thus provided the starting point for these analyses.

10.5.1 The existing WHOQOL data sets

In order to carry out the analyses, three main data sets were used, briefly described as follows.

- “OLD100” (N = 4802) is the original data set collected with the pilot version of the WHOQOL and used to produce the WHOQOL-100. The data originate from 15 centres worldwide with just under 48% (N = 2300) of the data originating from European respondents. This is the only global data set that also includes separate Importance ratings for each of the facets for each respondent.
- “NEW100” (N = 7701) was formed from the field trial of the WHOQOL-100 in 19 different centres worldwide. Although the majority of the original pilot WHOQOL centres also provided field trial data, this data set includes data from a number of new centres (e.g. Brazil, China and Germany) that were not included in the original study. Approximately 39% (N = 3032) of the data originate in European centres. The majority of this data set, together with the data from the OLD100 data set, was used to produce the short form of the WHOQOL, the WHOQOL-BREF.
- “BREFNEW1” (N = 11 830) was produced from the field trials of the WHOQOL-BREF and has recently been analysed to produce reliability and validity information on the WHOQOL-BREF. The data are therefore based on the 26 items of the WHOQOL-BREF with approximately 61% (N = 7251) originating from European centres. Data were available from a total of 32 different centres worldwide.

It should be noted that the choice of potential EUROHIS items was restricted to the 26 WHOQOL-BREF items rather than broadened to the WHOQOL-100 items. This restriction maximizes the use of the existing data sets and ensures future comparability between EUROHIS and WHOQOL-BREF data sets. The initial analysis was carried out in seven steps, as explained in the following section.

10.5.2 Initial analysis

STEP 1 – FREQUENCY ANALYSES

Item response frequency problems were analysed. As in previous published analyses of the WHOQOL, a frequency problem was flagged up where any two or more adjacent response categories had less than 10% of the total of the permissible responses. Again, following previous recommendations (Bullinger et al. 1996), items in which 50% or more of the centres showed frequency problems need to be highlighted for exclusion (item numbers refer to the position in the WHOQOL-BREF):

Possible frequency problem

F1 – How would you rate your quality of life?

Also of note

F11 – Are you able to accept your bodily appearance?

F13 – How available to you is the information that you need in your day-to-day life?

F20 – How satisfied are you with your personal relationships?

F24 – How satisfied are you with your access to health services?

STEP 2 – RELIABILITY ANALYSES

Scale internal reliability or consistency analyses were carried out using Cronbach alpha. These analyses demonstrate, first, whether or not an item contributes to its predicted scale and, second, whether or not elimination of the item would improve the internal consistency of the scale. In these particular analyses, Cronbach alphas were calculated for the four-domain BREF solution. Six-domain solutions were not tested because one of the domains (spirituality) would be a single-item domain for which internal reliability analyses would not be possible.

To be consistent with previous WHOQOL analyses and with general recommendations in the literature, any item with a corrected item-total correlation with its own domain of <0.4 was flagged up for consideration as being problematic. The items with the highest numbers of reliability problems across centres are noted below.

For exclusion

F8 – How safe do you feel in your daily life?

Two additional items, F21 and F22, also showed reliability problems but the fact that they are from a domain with only three items makes the Cronbach alpha analysis less reliable.

When the three data sets were divided into European versus non-European groupings, there were no consistent problems at this level of analysis.

STEP 3 – MAP ANALYSES

The so-called MAP analysis is an abbreviation for an adaptation of the Multi-trait Analysis Program used in previous WHOQOL analyses and based on software used in the development of the SF-36. In this simpler use, the MAP analyses are merely used to highlight items that load more highly on other subscales than on their own predicted subscale (the latter being in the form of the corrected item-total correlation).

When summarized across the three data sets and for both global and European-only analyses, the following items (starting with the most problematic) showed MAP problems:

F8 – How safe do you feel in your daily life?

F10 – Do you have enough energy for everyday life?

F16 – How satisfied are you with your sleep?

That is, these items tended to load more highly on other domains than on their own predicted domains. A number of other items showed problems in one or two of the analyses, but do not need to be flagged up here.

It must be noted however that the cross-loading of items need not be such a problem where the aim, as in the present analyses, is to derive a set of representative items. It depends of course on how “representativeness” is defined. That is, if items are being chosen to best represent a particular domain, an item that loaded more highly on another domain would be a contraindicated choice. If, however, items were simply being chosen to best represent overall QOL (thereby ignoring lower levels of domain or facet structure), then it would be feasible for such an item to be considered for inclusion.

The question of interpretation of the MAP analyses raises the issue of the level at which the EUROHIS Group wishes the nominated items to be represented. This issue was initially debated for the choice of items when the WHOQOL-BREF was derived from the WHOQOL-100. It was eventually decided to maintain the facet structure through the selection of one item to represent each of the 24 facets in the WHOQOL-100, together with the inclusion of two general items. The parallel decision for the EUROHIS group would be to select at least one item to represent each of the four domains, together with the possible inclusion of one of the BREF general items. A number of subsequent analyses therefore tested out item choices that retained domain structure.

STEP 4 – PRELIMINARY CORRELATIONAL, REGRESSION AND FACTOR ANALYSES

The correlational analyses are the simplest way of exploring the contribution of each item to its own domain. The analyses would suggest that the following are of particular importance:

- Domain 1 – F18, F17
- Domain 2 – F19, F6, F5
- Domain 3 – F20, F21
- Domain 4 – F12, F23, F25.

The exploratory factor analyses carried out for each domain separately would suggest the following items to be of note because they load highly on something resembling the appropriate factors:

- Domain 1 – F18, F17, F10, F4
- Domain 2 – F19, F6, F5
- Domain 3 – F20, F21
- Domain 4 – F12, F13, F24.

Two sets of multiple regression analyses were carried out. In the first set of analyses, items were tested in stepwise multiple regression to see how they contributed to their own total domain scores. The following items were of note:

- Domain 1 – F18, F17, F10, F4
- Domain 2 – F19, F6, F5
- Domain 3 – F21, F20, F22
- Domain 4 – F12, F23.

In the second set of multiple regression analyses, individual items within domains were used to predict the *overall* QOL score as assessed by the sum of the two general items included in the BREF (F1 and F2). From these analyses the following items were of note:

Domain 1 – F17, F10

Domain 2 – F19, F5, F26

Domain 3 – F20, F22

Domain 4 – F23, F8, F14, F12.

It is clear that these analyses begin to point to certain items as being more important statistically in their contribution to the different domains.

STEP 5 – CLUSTER ANALYSES

The hierarchical cluster analyses that were carried out with the different data sets were not easily interpretable, despite attempting several different cluster analysis approaches.

STEP 6 – MULTIDIMENSIONAL SCALING (MDS)

The SPSS ALSCAL program was used to carry out multidimensional scaling. A Euclidean distance, interval metric, replicated MDS approach was taken, which compared solutions for the European versus non-European centres. Examination of 2-D and 3-D solutions across the three data sets suggested that the following items were most central to the solutions examined:

Domain 1 – F10, F17, F18, F4

Domain 2 – F19

Domain 3 – (too few variables)

Domain 4 – F13, F24.

Although the patterns for Domains 1 and 2 seem reasonably consistent with the previous analyses, the patterns obtained for Domain 4 seemed distinct, though it is unclear why these differences occurred.

STEP 7 – ANALYSIS OF THE FACET IMPORTANCE DATA

Facet importance data were available only for the first data set collected with the pilot WHOQOL and used to derive the WHOQOL-100. Nevertheless, it was considered extremely useful to be able to provide some simple analyses of these data in order to inform the choice of items for the EUROHIS project. The facets were therefore simply ranked according to their mean importance ratings within each domain, with identical rankings across global and European centres as follows (beginning with the most important facet first):

Domain 1 – F17, F10, F15

Domain 2 – F5, F7, F19

Domain 3 – F20, F22, F21

Domain 4 – F23, F8, F12.

There are a couple of additional comments about the rank importance data. First, item F17 (How satisfied are you with your ability to perform your daily living activities?) was the item from the facet ranked most important in both the global and the European data sets. Second, item F21 (How satisfied are you with your sex life?) was the item from the facet ranked as least important in the data sets.

SUMMARY OF THE INITIAL ANALYSES

At this point it was considered possible that a summary of the analyses thus far could be used to highlight the “best” items so that a more restricted set of items would be taken forward to the next set of analyses. A particular subset of items could be formed based on the following criteria:

- the domain structure of the BREF should be retained in the EUROHIS items (for example, see if there are one or two items in each domain that can “best” represent that domain);
- an item must be rated as important within its domain;
- the item should appear as significant in most if not all of the prior analyses;
- the item should preferably have no frequency or reliability problems; and
- MAP analysis problems can probably be downplayed for the current purpose.

On the basis of these criteria, the following items were taken forward for the first set of further analyses.

General – F2: How satisfied are you with your health?

Domain 1 (Physical):

F10: Do you have enough energy for everyday life?

F17: How satisfied are you with your ability to perform your daily living activities?

Domain 2 (Psychological):

F5: How much do you enjoy life?

F19: How satisfied are you with yourself?

Domain 3 (Social):

F20: How satisfied are you with your personal relationships?

F21: How satisfied are you with your sex life?

Domain 4 (Environment):

F12: Have you enough money to meet your needs?

F23: How satisfied are you with the conditions of your living place?

10.5.3 Confirmatory factor analyses and Rasch analyses

The purpose of the confirmatory factor analyses (CFA) (carried out using the EQS program) and the Rasch analyses (carried out using the WINMIRA 32 program) was to help select from the restricted set of nine items those that might best reflect the domain and overall structure of the WHOQOL.

The purpose of the CFA was, first, to test which combinations, if any, of one item per domain plus one general item would show good fit indices for a single latent trait solution. Similarly, the Rasch analyses were also used to test which items were and were not consistent with a single underlying latent trait. It must be emphasized, therefore, that this use of the CFA and Rasch analyses is different to how these analyses have been used previously in the analysis of the WHOQOL. Whereas previous analyses have primarily been used to test the conceptual structure of the WHOQOL (e.g. the 4-domain and 6-

domain hierarchical models), the current analyses were primarily used as a method of selection of possible items for the EUROHIS project.

Summary of the CFA

In fact, the CFA showed that the combination of the general item (F2) with all combinations of four additional items (one per domain) showed satisfactory fit indices with, for example, all CFI² values greater than 0.9 (the normal agreed threshold value for an acceptable level of fit). That is, the CFA results showed that any combination of items selected along these lines could be adequately represented by a single factor or underlying trait; these findings were consistent across all data sets and were consistent across all European-only data analyses.

Summary of the Rasch analyses

The Rasch analyses were used as a way of eliminating items from the nine-item set but with the stated restriction that each domain should be represented. Within the Rasch analysis program used, WINMIRA 32, a significant value for the so-called Q-index (which is a class-specific item-fit measure for Rasch models) indicates that a particular item does not produce the expected pattern for a single underlying trait. Analysis of just the European data from each of the three data sets showed the following order of elimination of items from the data sets. That is, at the first step the item with the poorest fit (i.e. highest Q-index) was eliminated, the analysis was then repeated and the item with the poorest fit for the next step eliminated, and so on, while observing the restrictions noted above.

	<u>OLD100</u>	<u>NEW100</u>	<u>BREF</u>
Step 1	F12	F5	F12
Step 2	F5	F21	F21
Step 3	F20	F12	F5
Step 4			F10

The Rasch analyses would therefore support the elimination of F12, F5, F21 and possibly F10 (rather than F17) from Domain 1.

The inclusion of the Rasch analyses in the decision-making for the item choice would lead to the following five items being proposed:

F2: How satisfied are you with your health?

F17: How satisfied are you with your ability to perform your daily living activities?

F19: How satisfied are you with yourself?

F20: How satisfied are you with your personal relationships?

F23: How satisfied are you with the conditions of your living place?

² The CFI (Comparative Fit Index) tests the overall fit of the empirical model parameters with estimated model parameters.

Further comments and analyses

- The set of five items proposed conceptually cover a very good range of facets of QOL, including self-esteem, social support, home environment, daily living activities and general health. The only, perhaps minor, qualm is that all of the items are “satisfaction items”, although perhaps this is no surprise given that seven of the nine items taken forward for further analysis were such items.
- A CFA for the proposed five items has a CFI = 0.923 (chi-square = 1232.97) in a replicated test across the three European data sets that tests for a single factor solution.
- Cronbach alpha reliability analyses for the five items gave values of 0.72, 0.77 and 0.76 in the OLD100, NEW100 and BREF European data sets, respectively.
- Correlations between the summary score for the five items and the summary domain scores for the full BREF items in the three European data sets gave a range of correlations from 0.63 (with the OLD100 Domain 4 summary score) to 0.79 (with the BREF Domain 2 summary score). Given that 8 out of 12 of the correlations were at 0.70 or better, this suggests that there is a good overlap between the five proposed EUROHIS items and the full BREF domain scores.

10.5.4 Subsequent analyses

In commenting on the analyses conducted thus far, the QOL indicator group raised a number of issues for further discussion, as below. This led to a further set of analyses being conducted.

Point 1. Some of the results will depend on the composition of the samples, with sick/well status being particularly important.

Summary tables for the three data sets in relation to sick/well status were reviewed. The main message to take from these tables is that the original OLD100 data set is different to the other two because the sampling frame was set at 250 ill persons versus 50 well persons. In contrast, the NEW100 and BREFNEW data sets probably reflect a 50:50 sick/well composition, although the substantial amount of missing values makes this a guesstimate. Nevertheless, the greater similarity in the analyses for these data sets may simply reflect the fact that the OLD100 data set has a substantially different composition.

Point 2. Would the multidimensional scaling analyses have been more informative if all items had been included together, rather than being carried out at the domain level?

These analyses were, as a result, carried out for 2-D and 3-D solutions across the three data sets. If nothing else, these analyses flag up the centrality of the F1 item (How would you rate your quality of life?) and provide good justification for adding it back into the EQS and Rasch analyses (despite the problems with its frequency characteristics in the first set of analyses).

Point 3. The Rasch analyses were disappointing because they led to the choice of (health) satisfaction items only.

First of all, the Rasch analyses were reported as follows. Beginning with, for example, nine items, these would be tested to see if they corresponded to the Rasch model of a single underlying trait. The item that least corresponded to the Rasch model (as shown

by a high positive value for the Q-index or standardized Z_q value) was then dropped and the analysis repeated for eight items, and so on, until either all the remaining items complied with the Rasch model or no Rasch model was ever found to fit.

Second, it was disappointing to find that only “satisfaction” items seemed to comply with the Rasch model. This may simply reflect the linguistic overlap between phrasing of items (a point that Rex Billington has often made in his discussion of the WHOQOL), but it may also reflect the problem of including the general item “How satisfied are you with your health?”

Further Rasch analyses were therefore carried out that took the other general item “How would you rate your quality of life?” as their starting point. The results, however, were problematic across the three data sets, with no Rasch model fitting the OLD100 data set. Perhaps the most sensible view now is that the WHOQOL was never designed to have the properties of a Rasch-type scale, so one should take less notice of the Rasch analyses than of the other multivariate analyses. The assumptions that underlie the Rasch model may not be applicable to the present data – the model assumes unidimensionality, i.e. it focuses on a single parameter (that of the location of the items on a scale) and its values depend on a particular reference group.

Outcome of subsequent analyses

Taking as the starting point for the further analyses the inclusion of the other general item, then the following possible five items appeared to be a good alternative to those proposed previously:

- F1: How would you rate your quality of life?
- F10: Do you have enough energy for everyday life?
- F19: How satisfied are you with yourself?
- F20: How satisfied are you with your personal relationships?
- F12: Have you enough money to meet your needs?

10.5.5 Summary of results from analyses

The optimal way forward from the analyses that were conducted appeared to be to take the complete set of eight items that were generated from the two different sets of analyses and include them in the pilot EUROHIS survey. These data would then provide further information about the items from representative national populations, in addition to looking at their performance and properties in relation to other EUROHIS instruments. The eight items were as follows:

- F1: How would you rate your quality of life?
- F2: How satisfied are you with your health?
- F10: Do you have enough energy for everyday life?
- F17: How satisfied are you with your ability to perform your daily living activities?
- F19: How satisfied are you with yourself?
- F20: How satisfied are you with your personal relationships?
- F12: Have you enough money to meet your needs?
- F23: How satisfied are you with the conditions of your living place?

10.6 Translation of items

The existence of a wide range of language-specific versions of the WHOQOL reduced the need for new translations of the proposed EUROHIS survey items. Nevertheless, the fact that a number of pre-testing centres were willing to include the QOL indicator in their pre-testing of other indicators provided an opportunity for a check on existing translations of items, the production of translations into new language forms, and the collection of qualitative and quantitative feedback on the performance of the proposed items.

10.7 EUROHIS pre-testing 2000

A number of centres in western and eastern Europe collected responses on a subset of the draft EUROHIS indicators that included the QOL indicator. The feedback was typically based on a small sample of 20 or so individuals from whom both quantitative and qualitative information was collected. In general, this feedback was positive about the use of the eight-item QOL instrument, but several interesting points were raised.

- Specific phrases in some of the items do not have exact equivalents or can lead to ambiguous interpretations in some languages, for example, phrases such as “living conditions”, “daily living activities” and “living place”. The key point about translation in the WHOQOL methodology is that literal translation is not desirable, but rather the use of semantic equivalents and language-specific idiom should be made whenever appropriate.
- At least one centre reported that the term “quality of life” itself is not a concept that is readily understood within that culture. Again, as with the previous point, centres should aim for a semantically equivalent concept within that language even if the phrase “quality of life” itself has no direct translation.
- When “official” translations of the WHOQOL already exist in a particular language, these must be used rather than new translations, otherwise the data collected with the WHOQOL-BREF and with the EUROHIS-QOL might not be comparable for that culture.

These issues, together with the discussions that took place at the mid-term review meeting in Copenhagen on 2–4 November 2000, provided the following proposal for the draft instrument to be included in field-testing during 2001. It was considered that the measurement of QOL could be conceptualized at three different levels:

- *First level.* At the absolute minimum there should be at least one general question included:
- F1: How would you rate your quality of life?
- *Second level.* The eight items would provide a useful overview of the relevant domains of QOL for many health surveys.
- *Third level.* Where feasible, all the main 26 items from the WHOQOL-BREF could be included in a national survey, especially where QOL was intended to be a key indicator.

In summary, QOL provides an essential indicator for health surveys, but it is feasible to assess it at several different levels of detail – for example, as a single overview question, as a short set of eight items that could be combined to form an overall scale, or as an established scale that is already in widespread use in a range of studies.

10.8 EUROHIS field-testing 2001

The draft QOL instrument was field-tested alongside the other seven EUROHIS draft instruments. The results presented here are from France, Germany and the United Kingdom, where 400 telephone interviews were conducted in each country. Data were also collected on a range of supplementary socio-demographic variables.

The preliminary analyses of the individual items showed that there were very low rates of missing values owing, for example, to refusal to answer the question or to non-applicability. Across the three samples, almost all missing values were less than 1%, apart from Item 5 (How satisfied are you with yourself?) for the French sample, which had a missing value rate of 1.8%.

The item frequency distributions across the five response categories for each item showed that there was some skewing of the data to the positive end of the response scales. Problematic item distributions for WHOQOL items have previously been defined as less than 10% of the total responses for two or more adjacent categories (WHOQOL Group 1998). Only one of the items, Item 2 (How satisfied are you with your health?) had a good item distribution for all three samples with, in addition, Item 4 and Item 7 being well distributed in the British and French samples, respectively. In terms of population surveys therefore, it is interesting that the best distributed item is the health-related QOL item. The other items show that in western European representative populations there will inevitably be some skewing of the items towards the positive or “more satisfied” end of the scale.

A comparison of mean QOL item values across the three samples showed that there was only one item, Item 2, on which the three samples did not differ. The pattern of differences was very variable across the other items, and there was only one item, Item 3 (Do you have enough energy for everyday life?), for which the comparisons were all significant (United Kingdom mean = 3.9; France mean = 4.1; Germany mean = 4.3). In terms of the total score for all eight QOL items, the French sample scored significantly lower (mean = 30.9) than the British (mean = 32.0) and German (mean = 32.6) samples, which did not differ significantly from each other.

Analyses of the scale performance showed satisfactory Cronbach alpha values for internal consistency in all three centres (France = 0.734, Germany = 0.795, United Kingdom = 0.795), with the combined value therefore being satisfactory across the three samples (alpha = 0.779).

The final set of analyses provide a very preliminary look at how QOL might be related to some of the other indicators that were included in the data collected from the three samples. Examination of zero-order correlations between some of the socio-demographic and other indicator variables with the summated QOL score (formed by a simple summation of scores on the eight items, with higher scores indicating better QOL) showed a number of significant correlations. The respondent's overall rating of health correlated significantly with QOL ($r = -0.511$, $P < 0.001$), as did overall physical activity ($r = 0.228$, $P < 0.001$), total number of physical symptoms ($r = -0.423$, $P < 0.001$) and total psychological distress ($r = 0.541$, $P < 0.001$). These variables were then included in a number of regression equations to predict overall QOL, together with a number of dichotomized socio-demographic variables that included age (“young” versus “old”), marital status (“married” versus “the rest”), gender and employment status (“employed” versus “the rest”). The contributions to overall QOL of these eight predictor variables were then examined in a number of multiple regression equations. The best final equation for these variables is shown in Table 10.1.

Table 10.1. Multiple regression analysis: contribution of predictor variables to overall QOL

Variable	Beta	Sig. T
Age (old vs. young)	0.111	0.012
Marital status	0.107	0.002
Employment status	0.053	0.087
Self-reported health	0.363	0.001
Total physical symptoms	0.120	0.001
Physical activity level	0.107	0.001
Mental health	0.356	0.001

[Mult. $R = 0.687$; $R^2 = 0.472$; Adj. $R^2 = 0.467$]

Table 10.1 shows that seven out of the eight variables make a significant contribution using this particular combination of variables. That is, all of the variables apart from gender contribute, although the effect of employment status is rather marginal. Interestingly, there appears to be something of an interaction between age and employment status in the sense that, although overall QOL tends to improve with age, this is moderated by the loss of employment for older adults, which counteracts the overall age improvement. The other variables in the equation show clear and separate contributions to overall QOL from self-reported health status, total physical symptoms on the physical symptom indicator, total psychological distress on the mental health indicator, and overall physical activity level on the physical activity indicator. These analyses are of course only preliminary and they await, for example, the derivation of the best summary measures for indicators such as alcohol consumption, use of medicines and use of preventive health care. Nevertheless, they offer an initial step for examining the interaction between these health indicators in representative population surveys throughout Europe and other parts of the world.

Further exploratory analyses of the QOL data and how they relate to physical and mental health in both western and eastern European countries have been undertaken. These analyses suggest that QOL has a significant effect on both physical and mental health outcomes, and further emphasize the need for inclusion of a QOL indicator in national health surveys.

10.9 Conclusions

The different conceptual and empirical analyses that have been presented in this chapter point towards a number of important conclusions. First, it is clear that QOL is a key health indicator that should be routinely assessed in health surveys. Second, there is currently a range of instruments – both standardized and unstandardized – that have been used in health surveys to date. Our own analyses point to the strengths of the WHOQOL set of measures and suggest three levels of increasing detail at which QOL can be assessed:

- as a minimum, the single item “How would you rate your quality of life?”;
- an intermediate level that consists of eight items (see Appendix 10.1); and
- a detailed level that consists of the 26-item WHOQOL-BREF (see Appendix 10.2), which can provide scores for four key QOL domains in addition to an overall index.

The exact choice of which of these levels is used to assess QOL will then depend on the purpose and aims of the health survey, together with the level of detail at which it is desired to assess QOL.

References

- BECH, P. (1996) *The Bech, Hamilton and Zung Scales for mood disorders: screening and listening*, 2nd ed. Berlin, Springer.
- BECH, P. ET AL. (in press) Measuring well-being rather than absence of distress symptoms. A comparison of the SF-36 Mental Health subscale and the WHO-Five well-being scale. *International journal of methods in psychiatric research*.
- BULLINGER, M. ET AL. (1996) Creating and evaluating cross-cultural instruments. In: Spilker, B., ed. *Quality of life and pharmacoeconomics in clinical trials*, 2nd ed. Philadelphia, Lippincott-Raven, pp. 659–669.
- DUPUY, H.J. (1984) The Psychological General Well-Being (PGWB) Index. In: Wenger, N.K. et al., ed. *Assessment of quality of life in clinical trials of cardiovascular therapies*. New York, Le Jacq.
- POWER, M.J. ET AL. (1999) The World Health Organization WHOQOL-100: tests of the universality of quality of life in 15 different cultural groups worldwide. *Health psychology*, **18**: 495–505.
- SPIPKER, B. (1996) Introduction. In: Spilker, B., ed. *Quality of life and pharmacoeconomics in clinical trials*, 2nd ed. Philadelphia, Lippincott-Raven.
- 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.
- WHOQOL GROUP (1998) The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. *Social science and medicine*, **46**: 1569–1585.
- WORLD HEALTH ORGANIZATION (1948) *Constitution of the World Health Organization*. Geneva, World Health Organization, pp.1–2.
- WORLD HEALTH ORGANIZATION (2000) *Report of the second meeting of the Quality of Life Indicator Network, Barcelona, Spain, 6 November 1999*. Copenhagen, WHO Regional Office for Europe (document INFO 02 03 02).

Appendix 10.1

EUROHIS recommended common instrument for quality of life (field-tested in 2001)

Interviewer. This set of questions asks how you feel about your quality of life, health or other areas of your life. We ask that you think about your life in the past two weeks.

1. How would you rate your quality of life:

very poor,
poor,
neither poor nor good,
good, or
very good?

2. How satisfied are you with your health:

very dissatisfied,
dissatisfied,
neither satisfied nor dissatisfied,
satisfied, or
very satisfied?

3. Do you have enough energy for everyday life:

not at all,
a little,
moderately,
mostly, or
completely?

4. How satisfied are you with your ability to perform your daily living activities:

very dissatisfied,
dissatisfied,
neither satisfied nor dissatisfied,
satisfied, or
very satisfied?

5. How satisfied are you with yourself:

very dissatisfied,
dissatisfied,
neither satisfied nor dissatisfied,
satisfied, or
very satisfied?

6. How satisfied are you with your personal relationships:

very dissatisfied,
dissatisfied,
neither satisfied nor dissatisfied,
satisfied, or
very satisfied?

7. Have you enough money to meet your needs:

not at all,
a little,
moderately,
mostly, or
completely?

8. How satisfied are you with the conditions of your living place:

very dissatisfied,
dissatisfied,
neither satisfied nor dissatisfied,
satisfied, or
very satisfied?

Appendix 10.2 The WHOQOL-BREF

(The numbers in brackets refer to the question numbers in the WHOQOL-100)

		Very poor	Poor	Neither poor nor good	Good	Very good
1 (G1)	How would you rate your quality of life?	1	2	3	4	5

		Very dissatisfied	Dissatisfied	Neither satisfied nor dissatisfied	Satisfied	Very satisfied
2 (G4)	How satisfied are you with your health?	1	2	3	4	5

The following questions ask about **how much** you have experienced certain things in the last two weeks.

		Not at all	A little	A moderate amount	Very much	An extreme amount
3 (F1.4)	To what extent do you feel that (physical) pain prevents you from doing what you need to do?	1	2	3	4	5
4 (F11.3)	How much do you need any medical treatment to function in your daily life?	1	2	3	4	5
5 (F4.1)	How much do you enjoy life?	1	2	3	4	5
6 (F24.2)	To what extent do you feel your life to be meaningful?	1	2	3	4	5

		Not at all	A little	A moderate amount	Very much	Extremely
7 (F5.3)	How well are you able to concentrate?	1	2	3	4	5
8 (F16.1)	How safe do you feel in your daily life?	1	2	3	4	5
9 (F22.1)	How healthy is your physical environment?	1	2	3	4	5

The following questions ask about **how completely** you experienced or were able to do certain things in the last two weeks.

		Not at all	A little	Moderately	Mostly	Completely
10 (F2.1)	Do you have enough energy for everyday life?	1	2	3	4	5
11 (F7.1)	Are you able to accept your bodily appearance?	1	2	3	4	5
12 (F18.1)	Have you enough money to meet your needs?	1	2	3	4	5
13 (F20.1)	How available to you is the information that you need in your day-to-day life?	1	2	3	4	5
14 (F21.1)	To what extent do you have the opportunity for leisure activities?	1	2	3	4	5

		Very poor	Poor	Neither poor nor good	Good	Very good
15 (F9.1)	How well are you able to get around?	1	2	3	4	5

The following questions ask you to say how **good or satisfied** you have felt about various aspects of your life over the last two weeks.

		Very dissatisfied	Dissatisfied	Neither satisfied nor dissatisfied	Satisfied	Very satisfied
16 (F3.3)	How satisfied are you with your sleep?	1	2	3	4	5
17 (F10.3)	How satisfied are you with your ability to perform your daily living activities?	1	2	3	4	5
18 (F12.4)	How satisfied are you with your capacity for work?	1	2	3	4	5
19 (F6.3)	How satisfied are you with yourself?	1	2	3	4	5
20 (F13.3)	How satisfied are you with your personal relationships?	1	2	3	4	5
21 (F15.3)	How satisfied are you with your sex life?	1	2	3	4	5
22 (F14.4)	How satisfied are you with the support you get from your friends?	1	2	3	4	5
23 (F17.3)	How satisfied are you with the conditions of your living place?	1	2	3	4	5
24 (F19.3)	How satisfied are you with your access to health services?	1	2	3	4	5
25 (F23.3)	How satisfied are you with your transport?	1	2	3	4	5

The following question refers to **how often** you have felt or experienced certain things in the last two weeks.

		Never	Seldom	Quite often	Very often	Always
26 (F8.1)	How often do you have negative feelings such as blue mood, despair, anxiety, depression?	1	2	3	4	5

Chapter 11.

Conclusions

This book necessarily gives a brief account of the EUROHIS project and focuses on the main output – the recommended instruments for all eight health indicators. It is clear from the foregoing chapters, however, that an enormous amount of work has gone into the development of these instruments. Furthermore, the output from the project consists of more than just the instruments themselves. The very nature of the project has required that contacts be made across a broad spectrum of people within and between the various indicator groups, across different scientific disciplines and between different cultures. This has generated an extensive and invaluable network of contacts that has allowed the project to exploit the wide expertise that is available.

Such a broad project as EUROHIS inevitably generates other challenges besides the expected scientific ones. These arose not least because of the size of the project, which encompassed eight indicator network groups that included researchers from 11 European Union (EU) countries. While these groups needed to work independently to develop instruments of very different natures, they also had to work in a coordinated manner to follow the same general programme of work and to allow common input into pre-testing and later field-testing of the draft instruments. As a reflection of the large number of people and institutions involved and the complexity of the tasks, the project has required substantial effort to ensure this coordinated approach, both within and between the eight networks, and in the involvement of all EU and several non-EU countries in the development process, the pre-testing and the field-testing.

In summarizing the work that has gone into the EUROHIS project, there appear to be several main conclusions.

First, there is no doubt that the availability of cross-culturally comparable health data is of fundamental importance in the WHO European Region. Such data are a prerequisite for evidence-based public health policy that aims at good health and good health care for the whole population. The absence of such data makes it impossible to satisfactorily assess and monitor national and international levels of health. Even the countries with the most advanced information systems have gaps in the coverage of health status data for some population groups. This problem differs in magnitude and severity between countries, but is common to all and needs to be addressed.

In the light of the diversity of methods and instruments currently used in the WHO Member States, the scope of the EUROHIS project was purposely limited to the most relevant and the most widely measured indicators, with an emphasis on short common instruments for monitoring purposes rather than detailed and more comprehensive instruments. The initial premise of the project was that the harmonization of the inputs (i.e. the instruments that were currently used) would be an effective means to harmonize the outputs (i.e. cross-population health survey data). It was assumed that this would be possible both for data that were already available from past surveys and for data from future surveys. At the time of the mid-term review of the project, however, it was already clear that this assumption was not realistic for many of the indicators, i.e. physical activity, alcohol consumption, quality of life, mental health and preventive health services. It was revealed that the relatively new concepts measured by these indicators were rarely used in practice, and thus had not been measured in a sufficiently large number of surveys.

Furthermore, even for those facets of the concepts that probably had been adequately measured, it would be difficult to harmonize the data retrospectively. This assessment was later confirmed in the field-testing.

While a more comprehensive field-test of the draft instruments was not initially planned in the EUROHIS project, this proved to be useful in further refining the instruments and providing some evidence for their validity. The resulting qualitative data suggested that the instruments performed adequately in all countries. The quantitative evidence was more difficult to interpret. While the instruments performed well in all countries, the results from the telephone interviews in France, Germany and the United Kingdom cannot be reliably compared with those from the other field-testing countries, which were mainly limited to convenience sampling and face-to-face interviews because of resource constraints. These biases caused by the use of different field-test methods were probably further compounded by translation differences. Despite an attempt to produce conceptually equivalent translations, it is not known whether the translations are indeed fully equivalent, and local modifications of the instruments were made in some countries to accommodate culturally specific attitudes, behaviour or health service structures. The quantitative results from the field-testing must therefore be used very cautiously. This in no way detracts from the project, as formal international comparisons of the field-test results were not a prerequisite for accepting the recommended instruments. Of course, further analyses of the field-test data would be an added bonus, on condition that the limitations of the data are taken into account and clearly described.

The results of the EUROHIS project confirm that the use of common instruments (i.e. harmonized inputs) is a prerequisite, but not a guarantee, for obtaining cross-culturally comparable health data (i.e. harmonized outputs). The recommended common instruments eliminate the unnecessary differences that arise from the instruments that are currently used, and encourage emphasis on the most relevant and feasible aspects in terms of cross-cultural comparisons. However, the project has also shown the limits of input harmonization. First, different cultures have different social norms of what is acceptable behaviour and what are acceptable responses to questions on health, thus creating biases within the data. Second, living conditions and the level of health service provision differ in the WHO European Region and reflect a distinct east–west gradient of wealth, as measured for example by GDP per person. Such differences affect the relevance of the instruments in any given population.

Third, there appears to be different concepts of health (and thus differences in the way that people assess their own physical and/or mental health) in different parts of the European Region. Preliminary analysis of the EUROHIS data across all eight indicators and all field-testing countries (bearing in mind, however, the methodological differences that underlie these data) indicates that at least two population patterns of health conceptualization and self-evaluation may be discernible – one in the western and another in the eastern European countries (Schmidt et al. 2002).

Cross-cultural harmonization of inputs is by its very nature a long-term task. The most practical objective at present may be to harmonize indicators and instruments as far as possible while being explicit about the concepts that are not yet sufficiently comparable. The harmonization of concepts, which lies at the core of harmonization of indicators and instruments, is continuously shaped by the development of knowledge and technology, as well as health system reforms and population dynamics. The harmonization of indicators and instruments reflects, therefore, the underlying process of cross-cultural amalgamation, which has accelerated noticeably in the past decade. These issues raise the need for an international clearing-house to ensure continuity and cost-effectiveness of the harmonization work being undertaken by numerous teams in a wide range of countries. This clearing-house should focus on surveys for health policy and health monitoring, rather

than on pure scientific research and the development of measurement instruments. Practical expertise needs to be built up in Europe for the harmonization of data from health surveys that employ harmonized instruments. The description in this book of the development of a common instrument for alcohol consumption illustrates the kind of subject knowledge and epidemiological skills that may be required to judge the comparability of the results that are produced by a harmonized instrument. This example is likely to be relevant for all indicators that require self-reporting of health-related behaviour. Considerable expertise is already available in different locations in Europe, which should be linked into formal networks and managed dynamically. Relevant guidelines and methodology for harmonization need to be developed and agreed, while we also need a better understanding of cultural differences in the concepts of health and how these affect respondents' answers in interview surveys.

The EUROHIS project has focused on the comparability of health surveys and has uncovered some of the problems that arise when attempts are made to compare data from different countries. Of course, similar problems of cross-cultural comparability also arise when comparing data from other sources, such as health registers, and are just as complex and difficult to tackle. One of the main advantages of health surveys, however, is the relatively low cost and time involved in adapting them according to the needs of health policy-makers.

The output from the EUROHIS project provides researchers and policy-makers with a set of recommended common instruments for the measurement of eight health indicators. The development of these instruments has required careful consideration of relevant common concepts and how these should be defined and operationalized. The instruments are freely available for use by all countries, with the aim of enhancing national health information systems and facilitating cross-national comparisons of health data.

Reference

SCHMIDT, S. ET AL. (2002) *Cross-cultural analysis of relationships of health indicators across Europe: first results based on the EUROHIS project*. Copenhagen, WHO Regional Office for Europe (document EUR/02/5041391).

ANNEX 1.

EUROHIS INDICATOR NETWORKS

INDICATOR 1

Chronic Physical Conditions

Principal Investigator

Viviana Egidi
Central Director, Population and Territory Direction
Italian National Statistical Institute (ISTAT)
Via Adolfo Ravà 150
00148 Rome, Italy

Active Participants

Vittoria Buratta, Italian National Statistical Institute (ISTAT), Rome, Italy
Agnes de Bruin, Statistics Netherlands, Voorburg, Netherlands
Luisa Frova, Italian National Statistical Institute (ISTAT), Rome, Italy
Lidia Gargiulo, Italian National Statistical Institute (ISTAT), Rome, Italy
Emilio Gianicolo, Italian National Statistical Institute (ISTAT), Rome, Italy
Rosa Gispert, Departament de Sanitat i Seguretat Social, Barcelona, Spain
Markku Heliövaara, National Public Health Institute (KTL), Helsinki, Finland
Howard Meltzer, Office for National Statistics, London, United Kingdom
Sabrina Prati, Italian National Statistical Institute (ISTAT), Rome, Italy
Luciana Quattrocioni, Italian National Statistical Institute (ISTAT), Rome, Italy
Jean-Marie Robine, Institut national de la Santé et de la Recherche médicale, Montpellier, France

INDICATOR 2. Mental Health

Principal Investigator (from November 2000)

Howard Meltzer
Principal Social Survey Officer
Social Survey Division
Room D1/22
National Statistics
1 Drummond Gate
London SW1V 2QQ, United Kingdom

(until May 2000)

Martti Heikkinen, National Public Health Institute (KTL), Helsinki, Finland

Active Participants

Jordi Alonso, Institut Municipal d'Investigacio Medica (IMIM), Barcelona, Spain
John H. Henderson, Mental Health Europe, Haddington, Scotland, United Kingdom
Rachel Jenkins, Institute of Psychiatry, London, United Kingdom
Viviane Kovess, Mutuelle générale de l'Education nationale (MGEN), Paris, France
Christine Kuehner, Central Institute of Mental Health, Mannheim, Germany
Hans Ormel, University of Groningen, Groningen, Netherlands
Jean-Marie Robine, Institut national de la Santé et de la Recherche médicale, Montpellier, France

INDICATOR 3. Alcohol Consumption

Principal Investigator (until autumn 2001)

Jussi Simpura
Alcohol and Drug Research Unit
National Research and Development Centre for Welfare and Health (STAKES)
P.O. Box 220
FIN-00531 Helsinki, Finland

(Jussi Simpura now works at Statistics Finland, Helsinki)

Active Participants

Francesco Cipriani, Osservatorio Epidemiologico dell'Agenzia Regionale di Sanità della Regione Toscana, Prato, Italy
Ronald Knibbe, Universiteit Maastricht, Maastricht, Netherlands
Ludwig Kraus, Institut für Therapieforschung, Munich, Germany

INDICATOR 4. Physical Activity

Principal Investigator

Pekka Oja
Scientific Director
UKK Institute for Health Promotion Research
Kaupinpuistoinkatu 1
FIN-33501 Tampere, Finland

Active Participants

Peter Fentem, Emeritus Professor of Stroke Medicine, Great Chesterford, United Kingdom
José Antonio Ribeiro Maia, Laboratório de Cineantropometria, Porto, Portugal
Gert Mensink, Robert Koch-Institut, Berlin, Germany
Willem van Mechelen, Institute for Research in Extramural Medicine, VU University
Medical Center, Amsterdam, Netherlands

INDICATOR 5. Use of Curative Medical Services

Principal Investigator (from 01 November 2000)

Agnes de Bruin
Statistics Netherlands
Prinses Beatrixlaan 428
2273 AZ Voorburg, Netherlands

(until 31 October 2000)

Jaap van den Berg, Statistics Netherlands, Heerlen, Netherlands

Active Participants

Sisko Arinen, National Public Health Institute (KTL), Helsinki, Finland
Paola Baroncini, Italian National Statistical Institute (ISTAT), Rome, Italy
Vittoria Buratta, Italian National Statistical Institute (ISTAT), Rome, Italy
Milagros Garcia Barbero, WHO European Office for Integrated Health Care Services,
Barcelona, Spain
Lidia Gargiulo, Italian National Statistical Institute (ISTAT), Rome, Italy
Herman van Oyen, Institute of Public Health, Brussels, Belgium

Methodological research and analysis

Dirkjan Beukenhorst, Statistics Netherlands, Heerlen, Netherlands

INDICATOR 6. Use of Medicines

Principal Investigator

Timo Klaukka
Research Department
The Social Insurance Institution (KELA)
Nordenskiöldinkatu 12
FIN-00101 Helsinki, Finland

Active Participants

Manfred S. Green, Chaim Sheba Medical Center, Gertner Institute, Tel Hashomer, Israel
Ebba Holme Hansen, Royal Danish School of Pharmacy, Copenhagen, Denmark
Jaana Martikainen, Social Insurance Institution (KELA), Helsinki, Finland
John Yfantopoulos, National and Capodistrian University of Athens, Athens, Greece

INDICATOR 7. Use of Preventive Health Care

Principal Investigator

Marijke de Kleijn-de Vrankrijker
TNO Institute of Preventive Health Care
Department of Public Health
Wassenaarseweg 56
NL-2333 AL Leiden, Netherlands

Active Participants

Paola Baroncini, Italian National Statistical Institute (ISTAT), Rome, Italy
Wim Davidse, TNO Institute of Preventive Health Care, Leiden, Netherlands
Viviana Egidi, Italian National Statistical Institute (ISTAT), Rome, Italy
Eberhard Greiser, Institut für Präventionsforschung und Sozialmedizin (BIPS), Bremen, Germany
Olli Impivaara, Social Insurance Institution (KELA), Turku, Finland
Jean Tafforeau, Institute of Public Health, Brussels, Belgium

INDICATOR 8. Quality of Life

Principal Investigator (from May 2000)

Mick Power
Department of Psychiatry
University Hospital of Edinburgh
Morningside Park
Edinburgh EH10 5HF, United Kingdom

(until May 2000)

Anatoliy Nosikov, Quality Information and Knowledge Base, Health Information and Evidence, WHO Regional Office for Europe, Copenhagen, Denmark

Active Participants

Per Bech, WHO collaborating centre in mental health, Frederiksborg General Hospital, Hillerød, Denmark

Jakob Bjørner, University of Copenhagen, Copenhagen, Denmark (has left the project and now works at the Institute of Occupational Medicine, Copenhagen)

Monika Bullinger, University Hospital of Hamburg Eppendorf, Hamburg, Germany

Jiri Holub, Institute of Health Information and Statistics, Prague, Czech Republic

Ramona Lucas Carasco, Barcelona, Spain

Suzanne M. Skevington, Department of Psychology, University of Bath, Bath, United Kingdom

ANNEX 2.

LIST OF EUROHIS PROJECT PARTICIPANTS

1. Participants Eligible for Funding by the European Commission

Legal name of organization	Institute of Public Health
City	Brussels
Country	Belgium
Scientific Official responsible for the project	Dr Jean Tafforeau
E-Mail	Jean.Tafforeau@iph.fgov.be
Scientific Official responsible for the project	Dr Johan van der Heyden
E-Mail	johan.vanderheyden@iph.fgov.be
Scientific Official responsible for the project	Professor Herman Jo van Oyen
E-Mail	Herman.Vanoyen@iph.fgov.be
<hr/>	
Legal name of organization	National Institute of Public Health
City	Copenhagen
Country	Denmark
Scientific Official responsible for the project	Dr N.K. Rasmussen
E-Mail	nkr@niph.dk
<hr/>	
Legal name of organization	Psychiatric Institute, Frederiksborg General Hospital
City	Hillerød
Country	Denmark
Scientific Official responsible for the project	Professor Per Bech
E-Mail	pebe@fa.dk
<hr/>	
Legal name of organization	Royal Danish School of Pharmacy
City	Copenhagen
Country	Denmark
Scientific Official responsible for the project	Professor Ebba Holme Hansen
E-Mail	ehh@dfk.dk
<hr/>	
Legal name of organization	None
City	Copenhagen
Country	Denmark
Scientific Official responsible for the project	Ms Joan F. Davidson
E-Mail	liebgottdavidson@get2net.dk
<hr/>	

Legal name of organization	National Public Health Institute (KTL)
City	Helsinki
Country	Finland
Scientific Official responsible for the project E-Mail	Ms Sisko Arinen sisko.arinen@ktl.fi
Scientific Official responsible for the project E-Mail	Professor Arpo Aromaa arpo.aromaa@ktl.fi
Scientific Official responsible for the project E-Mail	Dr Markku Heliövaara markku.heliovaara@ktl.fi
<hr/>	
Legal name of organization	STAKES - National Research and Development Centre for Welfare and Health
City	Helsinki
Country	Finland
Scientific Official responsible for the project E-Mail (present)	Professor Jussi Simpura (until autumn 2001) jussi.simpura@stat.fi
<hr/>	
Legal name of organization	Social Insurance Institution (KELA)
City	Turku
Country	Finland
Scientific Official responsible for the project E-Mail	Mr Olli Impivaara, Ph.D. olli.impivaara@kela.fi
City	Helsinki
Country	Finland
Scientific Official responsible for the project E-Mail	Dr Timo Klaukka timo.klaukka@kela.fi
Scientific Official responsible for the project E-Mail	Dr Jaana Martikainen jaana.martikainen@kela.fi
<hr/>	
Legal name of organization	UKK Institute for Health Promotion Research
City	Tampere
Country	Finland
Scientific Official responsible for the project E-Mail	Dr Pekka Oja ukpeoj@uta.fi
<hr/>	

Legal name of organization	Institut national de la Santé et de la Recherche médicale (INSERM)
City	Montpellier
Country	France
Scientific Official responsible for the project	Mr Jean-Marie Robine
E-Mail	robine@valdorel.fnclcc.fr
Legal name of organization	Ministère de l'Emploi et de la Solidarité
City	Paris
Country	France
Scientific Official responsible for the project	Mr Gérard Badéyan
E-Mail	gerard.badeyan@sante.gouv.fr
Legal name of organization	Mutuelle générale de l'Education nationale (MGEN)
City	Paris
Country	France
Scientific Official responsible for the project	Dr Viviane Kovess
E-Mail	vkovess@mgen.fr
Legal name of organization	None
City	Strasbourg
Country	France
Scientific Official responsible for the project	Dr Iain Bamforth
E-Mail	iainbamforth@wanadoo.fr
Legal name of organization	Central Institute of Mental Health
City	Mannheim
Country	Germany
Scientific Official responsible for the project	Dr Christine Kuehner
E-Mail	kuehner@as200.zi-mannheim.de
Legal name of organization	Institut für Präventionsforschung und Sozialmedizin (BIPS)
City	Bremen
Country	Germany
Scientific Official responsible for the project	Professor E. Greiser
E-Mail	greiser@bips.uni-bremen.de

Legal name of organization	Institut für Therapieforschung (IFT)
City	Munich
Country	Germany
Scientific Official responsible for the project	Mr Ludwig Kraus
E-Mail	kraus@ift.de
Legal name of organization	Robert Koch-Institut
City	Berlin
Country	Germany
Scientific Official responsible for the project	Dr Gert Mensink
E-Mail	mensinkg@rki.de
Scientific Official responsible for the project	Dr Thomas Ziese
E-Mail	zieset@rki.de
Legal name of organization	University Hospital of Hamburg Eppendorf
City	Hamburg
Country	Germany
Scientific Official responsible for the project	Professor Dr M. Bullinger
E-Mail	bullinger@uke.uni-hamburg.de
Scientific Official responsible for the project	Dr Silke Schmidt
E-Mail	sischmid@uke.uni-hamburg.de
Legal name of organization	National and Capodistrian University of Athens
City	Athens
Country	Greece
Scientific Official responsible for the project	Professor John Yfantopoulos
E-Mail	yfa@otenet.gr
Legal name of organization	Department of Health and Children
City	Dublin
Country	Ireland
Scientific Official responsible for the project	Mr H. Magee
E-Mail	hugh_magee@health.irlgov.ie
Legal name of organization	None
City	Dublin
Country	Ireland
Scientific Official responsible for the project	Ms Sylvia Yates
E-Mail	

Legal name of organization	Gertner Institute, Chaim Sheba Medical Center
City	Tel-Hashomer
Country	Israel
Scientific Official responsible for the project	Professor Manfred S. Green
E-Mail	m.green@icdc.health.gov.il
<hr/>	
Legal name of organization	Istituto Superiore di Sanità
City	Rome
Country	Italy
Scientific Official responsible for the project	Dr Giovanni de Girolamo
E-Mail	gdg@iss.it
<hr/>	
Legal name of organization	Italian National Statistical Institute (ISTAT)
City	Rome
Country	Italy
Scientific Official responsible for the project	Ms Paola Baroncini
E-Mail	baroncini@istat.it
Scientific Official responsible for the project	Dr Vittoria Buratta
E-Mail	buratta@istat.it
Scientific Official responsible for the project	Ms Viviana Egidi
E-Mail	egidi@istat.it
Scientific Official responsible for the project	Ms Luisa Frova
E-Mail	frova@istat.it
Scientific Official responsible for the project	Ms Lidia Gargiulo
E-Mail	gargiulo@istat.it
Scientific Official responsible for the project	Mr Emilio Gianicolo
E-Mail	emilio.gianicolo@istat.it
Scientific Official responsible for the project	Ms Sabrina Prati
E-Mail	prati@istat.it
Scientific Official responsible for the project	Ms Luciana Quattrociochi
E-Mail	quattrociochi@istat.it
<hr/>	

Legal name of organization	Osservatorio Epidemiologico dell'Agazia Regionale di Sanità della Regione Toscana
City	Prato
Country	Italy
Scientific Official responsible for the project E-Mail	Dr Francesco Cipriani fcipriani@usl4.toscana.it
Legal name of organization	NIPO Het Marktonderzoekinstituut
City	Amsterdam
Country	Netherlands
Scientific Official responsible for the project E-Mail	Mr R.N. Bos info@nipo.nl
Legal name of organization	Statistics Netherlands
City	Heerlen
Country	Netherlands
Scientific Official responsible for the project E-Mail	Dr Dirkjan J. Beukenhorst dbkt@cbs.nl
City	Voorburg
Country	Netherlands
Scientific Official responsible for the project E-Mail	Ms Agnes de Bruin abun@cbs.nl
Legal name of organization	TNO Institute of Preventive Health Care
City	Leiden
Country	Netherlands
Scientific Official responsible for the project E-Mail	Dr W. Davidse w.davidse@pg.tno.nl
Scientific Official responsible for the project E-Mail	Dr Marijke de Kleijn-de Vrankrijker mw.dekleijn@pg.tno.nl
Legal name of organization	University of Groningen
City	Groningen
Country	Netherlands
Scientific Official responsible for the project E-Mail	Dr Hans Ormel j.ormel@med.rug.nl

Legal name of organization	University of Maastricht
City	Maastricht
Country	Netherlands
Scientific Official responsible for the project	Dr Ronald Knibbe
E-Mail	r.knibbe@zw.unimaas.nl
Legal name of organization	VU University Medical Center, Institute for Research in Extramural Medicine
City	Amsterdam
Country	Netherlands
Scientific Official responsible for the project	Professor Willem van Mechelen
E-Mail	w.van_mechelen.emgo@med.vu.nl
Legal name of organization	Statistics Norway
City	Oslo
Country	Norway
Scientific Official responsible for the project	Ms Jorun Ramm
E-Mail	jrm@ssb.no
Legal name of organization	Laboratório de Cineantropometria
City	Porto
Country	Portugal
Scientific Official responsible for the project	Mr José Antonio Ribeiro Maia
E-Mail	jmaia@fcdef.up.pt
Legal name of organization	National Institute of Health
City	Lisbon
Country	Portugal
Scientific Official responsible for the project	Ms M. de Jesus A. Charrua Graca
E-Mail	ceb.onsa@insarj.pt
Legal name of organization	Institut Municipal d'Investigació Mèdica (IMIM)
City	Barcelona
Country	Spain
Scientific Official responsible for the project	Dr Jordi Alonso
E-Mail	jalonso@imim.es
Legal name of organization	Ministry of Health and Consumer Affairs
City	Madrid
Country	Spain
Scientific Official responsible for the project	Dr L. Biglino
E-Mail	lbiglino@msc.es

Legal name of organization	Departament de Sanitat i Seguretat Social, Generalitat de Catalunya
City	Barcelona
Country	Spain
Scientific Official responsible for the project	Ms Rosa Gispert
E-Mail	rgispert@dsss.scs.es
Legal name of organization	None
City	Barcelona
Country	Spain
Scientific Official responsible for the project	Dr Ramona Lucas Carasco
E-Mail	18350rlc@comb.es
Legal name of organization	Institute of Psychiatry
City	London
Country	United Kingdom
Scientific Official responsible for the project	Dr Rachel Jenkins
E-Mail	r.jenkins@iop.kcl.ac.uk
Legal name of organization	Office for National Statistics
City	London
Country	United Kingdom
Scientific Official responsible for the project	Dr H. Meltzer
E-Mail	howard.meltzer@ons.gov.uk
Legal name of organization	University of Bath, Department of Psychology
City	Bath
Country	United Kingdom
Scientific Official responsible for the project	Professor S. Skevington
E-Mail	s.m.skevington@bath.ac.uk
Legal name of organization	University Hospital of Edinburgh
City	Edinburgh
Country	United Kingdom
Scientific Official responsible for the project	Professor M. Power
E-Mail	mj@srv1.med.ed.ac.uk
Legal name of organization	None
City	Great Chesterford
Country	United Kingdom
Scientific Official responsible for the project	Dr Peter Fentem
E-Mail	phf@fentem.demon.co.uk
Legal name of organization	None
City	Haddington
Country	United Kingdom
Scientific Official responsible for the project	Dr John H. Henderson
E-Mail	john@mhech.freeserve.co.uk

2. Participants not Eligible for Funding by the European Commission

Legal name of organization	Research Institute of Lung Diseases
City	Baku
Country	Azerbaijan
Scientific Official responsible for the project	Dr Leyli Aliyeva
E-Mail	fma@who.baku.az (Liaison Officer)
Legal name of organization	Public Health Institute Republika Srpska
City	Banja Luka
Country	Bosnia and Herzegovina
Scientific Official responsible for the project	Dr Milorad Balaban
E-Mail	kab.moh@bih.net.ba
Legal name of organization	Federal Public Health Institute
City	Sarajevo
Country	Bosnia and Herzegovina
Scientific Official responsible for the project	Dr A. Pilav
E-Mail	idanap@hotmail.com
Legal name of organization	National Statistical Institute (NSI)
City	Sofia
Country	Bulgaria
Scientific Official responsible for the project	Ms Finka Denkova
E-Mail	fdenkova@nsi.bg
Legal name of organization	Croatian National Institute of Public Health
City	Zagreb
Country	Croatia
Scientific Official responsible for the project	Dr M. Kuzman
E-Mail	marina.kuzman@zg.tel.hr
Legal name of organization	Institute of Health Information and Statistics of the Czech Republic
City	Prague
Country	Czech Republic
Scientific Official responsible for the project	Dr Jiri Holub
E-Mail	holub@uzis.cz
Scientific Official responsible for the project	Dr Vlasta Mazankova
E-Mail	mazankova@uzis.cz

Legal name of organization	Ministry of Social Affairs of Estonia
City	Tallinn
Country	Estonia
Scientific Official responsible for the project	Ms M. Ruuge
E-Mail	mare.ruuge@sm.ee
Legal name of organization	Hungarian Central Statistical Office
City	Budapest
Country	Hungary
Scientific Official responsible for the project	Ms Eva Gardos
E-Mail	eva.gardos@office.ksh.hu
Legal name of organization	Ministry of Health
City	Budapest
Country	Hungary
Scientific Official responsible for the project	Dr Zoltan Voko
E-Mail	vokoz@elender.hu
Legal name of organization	Republic Centre of Continuing Education for Medical Specialists and Pharmacists
City	Bishkek
Country	Kyrgyzstan
Scientific Official responsible for the project	Ms Gulnara Yusupova
E-Mail	wholokgz@elcat.kg (Liaison Officer)
Legal name of organization	State Health Promotion Center
City	Riga
Country	Latvia
Scientific Official responsible for the project	Ms Iveta Pudule
E-Mail	iveta.vvc@parks.lv
Legal name of organization	Kaunas University of Medicine
City	Kaunas
Country	Lithuania
Scientific Official responsible for the project	Professor Vilius Grabauskas
E-Mail	vilgra@kmu.lt
Scientific Official responsible for the project	Ms Kristina Jureniene
E-Mail	krista@kmu.lt
Scientific Official responsible for the project	Dr Abdonas Tamosiunas
E-Mail	atamos@kmu.lt

Legal name of organization	Department of Health Information, Ministry of Health
City	Guardamangia
Country	Malta
Scientific Official responsible for the project	Dr R. Pace Asciak
E-Mail	renzo.pace-asciak@gov.mt
Legal name of organization	Centre of Public Health and Management
City	Chisinau
Country	Republic of Moldova
Scientific Official responsible for the project	Dr V. Sava
E-Mail	mam@moldnet.md
Legal name of organization	National Centre for Health Statistics and Information
City	Bucharest
Country	Romania
Scientific Official responsible for the project	Ms D. Deleanu
E-Mail	deleanu@ms.ro
Scientific Official responsible for the project	Dr Ioana Pertache
E-Mail	ipertach@ms.ro
Scientific Official responsible for the project	Dr D. Ursuleanu
E-Mail	dursulea@ms.ro
Legal name of organization	Semashko Institute, Russian Academy of Medical Sciences
City	Moscow
Country	Russian Federation
Scientific Official responsible for the project	Dr Vitaliy Belov
E-Mail	tmaximova@mail.ru (Professor Maximova's e-mail)
Scientific Official responsible for the project	Ms Olga Gaenko
E-Mail	tmaximova@mail.ru (Professor Maximova's e-mail)
Scientific Official responsible for the project	Professor T. M. Maximova
E-Mail	tmaximova@mail.ru
Legal name of organization	Trnava University
City	Bratislava
Country	Slovakia
Scientific Official responsible for the project	Dr Gabriel Gulis
E-Mail	gulissk@yahoo.com

Legal name of organization	Institute of Public Health of the Republic of Slovenia
City	Ljubljana
Country	Slovenia
Scientific Official responsible for the project E-Mail	Dr M. Macarol-Hiti metka.hiti@gov.si
Scientific Official responsible for the project E-Mail	Dr Andrej Marusic andrej.marusic@ivz-rs.si
Legal name of organization	Institute of Epidemiology, Biostatistics and Medical Informatics
City	Skopje
Country	The former Yugoslav Republic of Macedonia
Scientific Official responsible for the project E-Mail	Professor R. Isjanovska isjanovska@yahoo.com
Legal name of organization	Hacettepe University, Institute of Population Studies
City	Ankara
Country	Turkey
Scientific Official responsible for the project E-Mail	Dr A. Toros atoros@hacettepe.edu.tr
Legal name of organization	Ukrainian Institute of Public Health
City	Kiev
Country	Ukraine
Scientific Official responsible for the project E-Mail	Dr Irene Koveshnikova health@uiph.kiev.ua
Scientific Official responsible for the project E-Mail	Professor A. Nahorna health@uiph.kiev.ua



ISBN 1 58603 322 0 (IOS Press)
ISBN 4 274 90579 9 C3047 (Ohmsha)
ISSN 0929-6743