EUROPEAN COMMISSION
EUROSTAT
Directorate D: Single Market, Employment and Social statistics
Unit D-6: Health and food safety

OFFICE FOR NATIONAL STATISTICS

PARTNERSHIP ON HEALTH STATISTICS

Core group health and health related surveys (HIS core group)

Guidelines for the development and criteria for the adoption of Health Survey instruments

Institut Scientifique de Santé Publique
Instituto Nacional de Estadística
Kansanterveyslaitos
Folkhälsoinstitutet
National Public Health Institute
Istat
ROBERT KOCH INSTITUT

These guidelines were developed by Dr. Jean Tafforeau (coordinator), Montserrat Lopez Cobo, Hanna Tolonen, Christa Scheidt-Nave and Alessandra Tinto[1]

[1] The document was improved on the basis of the comments received from Arpo Aromaa (Finland), Herman Van Oyen (Belgium), Jozsef Vitrai (Hungary), Erzebet Stokker (Hungary)

Table of contents

Preface

Increasing attention has been given to the methodological requirements to be met in order to obtain comparability between data collected through Health Surveys in the different European Community member states. Two areas are concerned:

- content and validity of survey instruments
- survey procedures.

The aim of the present document is to propose guidelines concerning the development and the adoption of instruments to be used in population health surveys in the European Community member states (either at to national or at the European level). These guidelines can be used in order to follow-up and evaluate the projects in charge of instrument development financed by European Commission and can provide the criteria to decide whether a specific instrument can effectively be recommended for health surveys in Europe or not.

This is a generic document in the sense that it can be applied to any health survey instrument. Although most of the concepts used could probably also be applied to health examination surveys, the present guidelines are however for the moment more specifically focused on the health interview surveys.

Different steps are described here; it is suggested that experts who are developing or recommending instruments to be applied in population health survey in European Community member states may follow each of those steps to produce comprehensive documentation together with the instrument proposed.

It is also suggested that those experts who are using the recommended instruments should verify exactly that all the criteria proposed in the present guidelines are met before adopting a (new) instrument.

The guidelines are subdivided into five chapters:

- **Basic information**, outlining the background information to be provided
- **Development**, describing all the steps to be followed to design the instrument
- **Translation process**, with clear recommendations on how to ensure the comprehensive translation of the instrument from the source (English) to the target European languages
- **Quality evaluation** describ ing the minimal requirements concerning validation of the instrument
- **Implementation**, with a description of the content of the users'manual to be provided with the (new) instrument.

Each of those topics is briefly described here but the reader is also strongly recommended to study the corresponding information available in the scientific literature.

Definitions

The term “health survey” will be used here as a generic expression to label any kind of survey including health related topics.

The term “population survey” is used indicate surveys implemented on a representative sample of the population.

The term “Health Interview Survey” refers to a population survey specifically designed to investigate health topics; data collection is achieved by means of survey questionnaires without performing any kind of physical examination or biological testing.

The term “Health Examination Survey” refers to a population survey specifically designed to investigate health topics; data collection is achieved by means of survey questionnaires and by performing physical examination and/or biological testing.

The term “instrument” refers to one or several questions included in the survey questionnaire and aiming to investigate a specific subject matter (by example, smoking behaviour, self perceived health, …).

The term “measure” refers to the outcome of the instrument, that is the results obtained from one or several questions on a specific topic.

The term “module” refers to one or several instruments included in the survey questionnaire; together these instruments permit the investigation of a general area related to the health of the population such as health status, health behaviour, health care consumption, etc.

1. Introduction

(J. Tafforeau, Scientific Institute of Public Health, Brussels, Belgium)

Background

Health information and research has been defined by the 43rd World Health Assembly as a process for obtaining systematic knowledge and technology that can be used for improvement of the health of individuals or groups of population. Health information can thus be considered as one of the tools to be used for health promotion, disease prevention and health care management.

Due to the frequent lack of high quality and timely health data needed by the decision makers when designing public health programs, most of the European Community member states decided to develop tools aimed at gathering this information. Several countries (and most of the member states) facing the same situation have successfully responded to this problem by developing population health surveys. The pioneering countries, during the beginning of the eighties, were Canada, Denmark, the Netherlands and the United Kingdom. In these countries health surveys progressively became the necessary supplement to routine information systems in order to develop consistent public health policies.

Health interview surveys (HIS) provide useful information on health; they are based either on what people have been told (answering such a question as “have you ever been told by a doctor that you have diabetes?”) or on what they themselves perceive (answering such a question as “how is your health in general?”).

Some data for health monitoring can also be obtained from regular statistical sources and registers, e.g. hospital discharge or general practice registers. These data sources can provide an overview on morbidity and suggest hypotheses for further investigation (1). Comprehensive population register data are available only for some specific diseases, with differences as far as coverage and availability of data are concerned. In general, register data are valuable for the evaluation of health care services, but not sufficient for population health monitoring purposes (2) There are two main reasons for this: firstly, register data are subject to selection bias, as health service users differ from the total population. Secondly, estimates of prevalence are difficult to obtain, as the denominator remains unclear or must be approximated (i.e. total number of patients seen within a particular time period); in addition, the numerator is sometimes also questionable due to the lack of exhaustivity of the registration process.

Population health surveys can overcome much of the selection bias affecting register data, provided that the participation rate ot the survey is high in all population sub-groups. In addition, such surveys allow measurement of all the dimensions of health (as defined by WHO).

The added value of the population health surveys is the horizontal approach of the data collection: several types of information (health status, health determinants, personal characteristics, health consumption, etc ...) are collected simultaneously from the same persons. This makes it possible to produce a global picture of the health of the population and to identify priority areas. In addition, when the data are gathered periodically over time, changes in health and effects of health policies and interventions can be monitored.

The population health survey brings together the arguments for an increased investment in health promotion and prevention, and rationalisation of health care and expenditures. This information thus provides a powerful framework for a rational policy decision-making process (3). A provider’s perspective of health care will of course also be needed for a complete evaluation (HIS providing only the population user’s perspective), more particularly when investigating the quality of care.

On the other hand the results of health surveys have to be interpreted with caution as compared with more objective data coming from registers or services statistics. Selection bias may result from non-response due to those people who refuse to participate or could not be reached. As the data are collected in a sample of the population, statistical methods have to be applied taking into account the sampling design in order to interpret the result adequately. In addition, due to the relatively small sample size, health surveys are usually not suitable neither for monitoring rare events or diseases, nor for health monitoring in small geographical areas. For these purposes particular surveys targeted to special populations or applying small-area methodology are more appropriate.

Objectives of health surveys

The main objective of population health surveys is to provide a description of the health status of the population. The purpose is to obtain information on how people experience their health, to what extent they make use of health care facilities, and how they look after their health by adopting a certain life style or relying on preventive and other health services.

More specifically, the goals of health surveys can be summarized as follows:

- Identification of health problems
- Description of the health status and health needs of the population
- Measurement of the health status of the population
- Collection of data on health determinants
- Analysis of social (in)equality in health and access to health services
- Study of health consumption and its determinants as well as preventive care
- Study of possible trends in health status, lifestyle and health consumption of the population.

A health survey provides the channel through which such information can be obtained. On the basis of such surveys, assessing a large variety of personal, social and material characteristics, life habits and conditions, determinants for public health can be traced and monitored. The ultimate goal of a health survey is to provide an integrated instrument in decision-making when health policy is framed. This makes it possible to fix priorities in policy development and to monitor the progress of the health of the populations.

Content of health surveys

In population health surveys, data are collected by means of interviews in a representative sample of the population.

Most European member states conduct surveys exclusively oriented towards health (status and consumption) and its determinants. In some countries such as Germany and the UK (and more recently in the Netherlands with the “POLS”) there are multipurpose surveys with a specific module on health. However specific health surveys usually allow more in-depth and broader study of health-related issues than general surveys with health modules, as there is no competition with other areas of investigation (4).

Most health surveys are performed via interviews only (Health Interview Surveys – HIS). The potential advantages of examinations coupled with the interview must be analysed in detail due to the increased cost and complexity of such an investigation. Such health examination surveys (HES) are thus not discussed here.

Five main areas are usually considered in the conceptual framework of the Health Interview Surveys (HIS):

- Health Status
- Health behaviour
- Disease prevention
- Health consumption
- Health and society

These can be related to the classification proposed in the framework of the European Community Health Indicators – ECHI (5):

ECHI classification	HIS domains
1. Demographic and socio-economic factors 1.1 Population 1.2 Socio-economic factors	 Health and society
2. Health status 2.1 Mortality 2.2 Morbidity, disease-specific 2.3 Generic health status 2.4 Composite health status measures	 Health Status Health Status Health Status
3. Determinants of health 3.1 Personal and biological factors 3.2 Health behaviours 3.3 Living and working conditions	 Health behaviour Health and society
4. Health systems 4.1 Prevention, health protection and promotion 4.2 Health care resources 4.3 Health care utilisation 4.4 Health expenditures and financing 4.5 Health care quality/performance	 Disease prevention Health consumption Health and society

Health status

Attention to the health status of the population is necessary under the WHO definition of health and the global approach of health problems. Indeed, measuring health consumption only is no longer sufficient, and an instrument such as HIS is essential to complement the information usually collected by health care providers, registries and vital statistics.

HIS allows measurement of the health status of the population in general and not only in relation with specific health problems. This is referred to in the literature as the distinction between 'health status' and 'state of the health' (6).

Even if health is the prevailing subject of the survey and despite the positive approach of health recommended by WHO, most of the domains investigated in HIS will be concerned with ill-health and diseases. A positive conceptual framework was effectively considered when designing the HIS but it has unfortunately not been possible to implement the concepts due particularly to the lack of available instruments (7;8).

One of the main characteristics of the health survey comes from the fact that most of the information gathered is provided by the individuals themselves with all the potential subjectivity involved. Their experience and their sensitivity in relation to their own health status plays a major role. However it is possible to differentiate relatively more objective questions (height and weight for example) from purely subjective ones (self-perceived health). Most of the topics investigated in HIS lie between these two extremes.

Another basic concept of HIS is the different approach towards medically diagnosed diseases on one side and their consequences on the functional status of the individual on the other side. Here medical diagnoses refer to the declaration of the person answering the question "Has a doctor ever told you that you have ….?" without any objective verification by medical records.

The measurement of the health status of the population is mainly focused on chronic conditions; due to their long duration these conditions have a bigger impact on health expenditures and they represent a higher burden at the population level.

Not only the conditions are considered but also their impact on the functional status of the respondents: functionality and disability are thus also important problems that are investigated here.

Health behaviour

Lifestyles are intrinsic components of the daily lives of individual people. They are closely linked to the values and the priorities of each person, as well as to the opportunities and constraints inherent in culture and socio-economic status. Lifestyles are in fact shaped by the social acquirements and interpersonal interactions. It is thus wrong to believe that a specific behaviour is determined only by a simple personal decision to adopt or reject health-related life styles (deterministic approach).

Lifestyles are, however, health determinants: some aspects of daily life contribute to the preservation of a good state of health, the prevention of specific conditions and the improvement of psychological well-being. Equally, specific behaviours may be harmful to health especially if they are excessive or chronic.

Better lifestyles are the main potential source of improvement in the health of the population, probably to an even greater extent than medical progress. This is why health promotion is one of the most important components of the public health programmes especially the Health for All initiative of the WHO aimed at improving individual health-related behaviours.
It should be kept in mind that a reverse effect may also occur, for example people might give up smoking because of health problems, and not only to adopt a healthy behaviour.

For public health decision makers, as well as for the institutions in charge of the implementation of health-promotion programmes, it is essential to measure regularly the prevalence of specific health-related behaviours and their trends at population level and in specific population subgroups. See for example the Surf – NCD InfoBase of the World Health Organisation with estimates of national prevalence for each risk factor and Member State[2].

Such measurement is imperative for the evaluation of programmes and policies. Health surveys cannot be used to prove the relationship between a programme and a specific trend but they are however useful tools for the monitoring of health-related behaviours.

[2] http://www.who.int/ncd_surveillance/infobase

Disease prevention

The advantages of preventive medicine have become more and more apparent during the last 30 to 40 years. This new approach modifies deeply the way to solve problems such as infectious diseases (with immunisation programmes for example). Early disease detection has also become an essential component of preventive medicine with striking results as far as morbidity and mortality are concerned (9).

Public health policies have progressively been enlarged from the management of health care expenditures to the development of strategies aimed at improvement of the health of the population. Such an approach involves specific actions at the level of the biological factors, the physical and social environment, and the individual behaviour but also at the level of the health services in their curative and preventive components (10).

The WHO Health for All targets published in 1985 explicitly mention health promotion and diseases prevention programmes as priority programmes. From a conceptual point of view, three areas can be mentioned in the field of preventive medicine (10):

- Primary prevention: actions aimed at eradicating the cause of a disease in order to prevent emergence of new cases
- Secondary prevention: early detection and treatment of a specific disease before the appearance of the clinical symptoms and the complications
- Tertiary prevention: it is not strictly speaking prevention of diseases but rather trying to limit its consequences such as disability and chronic pain.

Some modules in the HIS investigate specific action in the primary and secondary prevention. Several methods have been used to select the priority actions in the area of preventive medicine. The frequency of the disease but also the importance of the problem at the individual and societal level, and the efficacy of the preventive methods are taken into account.

Health care consumption

Information on health care consumption is an essential part of the health information system in order to assign necessary resources to the population. This covers three main topics: ambulatory care, institutional care and medical drugs consumption.

Different methods are usually available to measure health care consumption: routine services statistics (including hospital discharge), health expenditures and their reimbursements by the social security system, and health surveys.

It is generally admitted that health services statistics are more reliable than information coming from health surveys. This is due to the recall bias as well as the lack of medical knowledge of the individuals participating in health surveys.

However, health surveys are the main source of information where data can be collected concurrently on different health related aspects, making it possible to analyse:

- the level of health consumption in correlation with several determinants such as health status, lifestyles or socio-demographic characteristics
- the relation between different types of health care use.

Health survey data permit the comparison of the health needs and health consumption and thus make it possible to explore the concepts of vertical and horizontal equity in health care.

Health surveys are also sometimes used to measure patient satisfaction.

Health and society

The concept of health has enlarged over time including progressively non-medical components and has become a social issue.
The health status and the social level of the individual are known to be closely linked: social status is a powerful determinant of health in the population. This can be studied through the accessibility to health care but also through the detailed analysis of the determinants of health inequalities.

In addition environmental (physical and social), as well as familial and professional parameters have to be considered, as well as income inequalities, social capital, social support , as they may interfere with the health of the population.

Health surveys are also used more and more frequently to investigate such problems as injuries, violence or gender inequalities. Ethnicity and geographical regions are important factors to be taken into account.

2. Basic Information on the phenomenon under study

(M. Lopez Cobo, National Institute for Statistics, Madrid, Spain)

Policy relevance and utility

The general aim of population health surveys is to acquire information about the state of the population with regard to health condition and its determinants, as far it is perceived. This knowledge is utilised to elaborate public health policies at the national and supranational levels, and to provide information to the population. Thus, any health survey or instrument to be developed needs to serve this purpose.

It is important to have clearly defined the main and specific objectives of the instrument to be developed. Only after we have defined what we want to know and how we want to get it, can we develop the instrument that provides us with the data to reach the predefined target.

In the process of selecting and designing an instrument, it is essential to justify the necessity of it, as well as to explain the future or expected use of the provided data. A list of potential users of the intended instrument could be provided.

In general, the utility of the data is not restricted to a single feature. A variety of factors can be involved, such as:

- the extent to which the predefined objectives of the survey and of a specific instrument have been reached.

- The possibility to compare the results at the geographical level (national and international) as well as the evolution over time.

- The possibility to compute the expected indicators, preferably those recommended internationally. Microdata are non-sense if these are not used to provide indicators, which summarise the information.

- The satisfaction of the largest number of possible users: local, national and supranational governments, which will benefit from the data in order to establish specific health policies (assistance, prevention, services, etc.). Non-Governmental Organisations, such as patients' organisations, may need to know the magnitude and geographical distribution of a disease or a disability, in order to better apply resources and maximise social benefits. Other probable data consumers are companies in charge of the supply of health products and services.

The utility of the data may arise from an urgent need for information about a particular problem. In that case, the introduction of a new instrument in a survey may make it possible to quantify the dimension of the problem and to measure the determinants; this will allow the planning of prevention strategies.

An example that combines political interest and production of internationally comparable indicators is the 2003 declaration of *the European Year of People With Disabilities,* and the ongoing activities in this context throughout the world and in particular in Europe.

In conclusion, the policy relevance and the utility of the instrument proposed have to be explained in detail in the manual that will be prepared for the potential users.

Justification of the inclusion of the module in the survey

The expert developing the instrument should justify why it should be included in a population survey (health or general survey). Is this the best source of data on the phenomenon under study? Are there any specific conditions to be fulfilled to include such an instrument in a population survey?

For example, the widespread instrument on self-perceived health, *"How is your health in general?"* is often included in population surveys, since this single question provides information about how people feel their own health, giving clues on unmet needs, health services, etc. In addition, interview surveys are the only way to collect such information from the population. Due to its simple wording, it may be included in both a general and a specific health survey. Recent studies (still to be published) have however shown a lack of intercultural comparability of this instrument.

On the other hand**,** an instrument to measure the prevalence of low birth weight is probably not adapted for inclusion in a population health survey. Indeed, due to the relatively small sample size of a generic survey, the number of new-borns included would be fairly small and would not allow computation of valid estimates. In this case other data sources such as birth registries would probably be preferable.

Description of the concept

For each instrument it should be explained <u>what</u> exactly will be measured.

The concept needs to be defined in the most complete and clear way, specifying what is meant to be included and what to be excluded in the study. It is usually necessary to complement this definition with other concepts related or included in an explicit or implicit way, also adding explanatory nuances: duration of the process, time reference indicating the period to which the instrument restricts, fulfilment of administrative conditions, etc.

For example:

A specific concept was adopted in 1992 entitled: *"health-related physical activity"* or *"health-enhancing physical activity"*. It is defined as any body movement produced by striated muscles leading to a significant increase of energy expenditure when compared with the rest status (Bouchard et al, 1994). This concept covers the entire spectrum of activities including leisure-time physical activities, exercises, competitive sports, occupational activities and daily tasks.

Interviewees will thus be asked if they performed any type of physical activity. Intensity as well as frequency of the effort will be taken into account. This will not be done through direct measurements but it will rather be based on the declaration on the individuals, with all the subjectivity that this may entail.

Another example:

The final report of the project "EuroREVES. Selection of a Coherent Set of Health Indicators", supported by the European Commission (11) refers to two different perspectives to tackle the concept of *mental health*. One is derived from a broad definition of health: *"Mental health is a positive sense of well-being; a belief in our own worth and the dignity and worth of others; the ability to deal with the inner world of thinking, feeling, managing life and taking risks; the ability to initiate, develop and sustain mutually satisfying personal relationships; the ability of the mind to heal itself after shock or stress"*. Detractors of this approach reject it as being impractical, and instead propose, *"mental health is the absence of mental disorder"*.

It is thus essential when developing an instrument to be used in health surveys, to clearly define the concept the instrument has been built on.

Description of the measure and the instrument

Once the meaning and coverage of the concept under study is appropriately delimited, it is necessary to describe the instrument selected in order to have the health issue measured. The range of measures that will be obtained must also be delimited, that is the results that will fulfil the requirements of information in the area being considered (Detailed information on criteria to be considered for the selection of the instrument is provided in chapter 3).

For example, in the context of the performance of activities of daily living, if the concept 'activity restriction' is defined regarding the guarantee of a minimal independence, the *Activities of Daily Living (ADL)* instrument would be considered to measure *'the restriction in activities which are considered essential to ensure the minimal independence in personal care (feeding, transfer, dress, etc.)'*. However, we could also measure a broader range of independence, '*that one which allow individuals to live alone in a private household'*, through the *Instrumental* Acti*vities of Daily Living (IADL)* instrument.

It is important also to explain what is the linkage or proximity between what it is really being measured and the concept defined *a priori*. For instance, in a module on chronic condition what it was aimed to measure is the prevalence of some chronic diseases but in interview surveys the respondent is asked if a doctor ever told him that he was suffering from some illnesses. This is thus only an indirect estimation.

Indicators

The International Institute for Sustainable Development (http://www.iisd.org/) gives a definition for 'indicator': "An indicator quantifies and simplifies phenomena and helps us understand complex realities. Indicators are aggregates of raw and processed data but they can be further aggregated to form complex indices." Thus, an indicator is a quantification of a measure.

Additional readings about the concept of health indicators and their role in the management of public health programmes can be found in the literature (12-17).

An Interesting website on indicators used in Canada is also available:

http://www.statcan.ca/english/freepub/82-221-XIE/free.htm.

More specifically, health indicators can be defined as quantitative measures chosen to reflect the health status of the population or to represent how well a health system is performing. By means of health indicators, the level and change in community health and in health system performance are judged.

The selected indicators must summarise the wide range of information given by the survey/instrument and fulfil its objectives. The instruments have to be such that they allow these indicators to be computed.

Several classifications of indicators exist:

- Classification of WHO's HFA by the Year 2000 strategy
- Classification by WHO for managerial process for national health development
- Classification for monitoring the European Regional strategy of WHO

- Classification of UN Statistical Office
- OECD list of health indicators
- …

Three levels of indicators are usually proposed:

1. Indicators associated with the health status of persons and populations in a given area
2. Indicators related to physical environmental conditions having a more or less direct bearing on the health status of the area under review
3. Indicators concerned with health services and activities directed to the improvement of health conditions

Since it is possible to determine an infinite number of indicators, users must make a selection of the information they need most. It is better to have a small number of good indicators than a large number that cannot be managed.

Indicators have to be compared with a standard in order to make a useful analysis of the indicators. Some of these standards are:

- Historic average of the indicator
- A standard carefully prepared by using work measurement methods
- Indicators from other sources
- Standards to be accomplished (targets)
- Statistical trends

Information on the indicators that the instrument will allow us to measure have to be provided in the manual that will be prepared for the potential users.

3. Development of the instrument

(J. Tafforeau, Scientific Institute of Public Health, Brussels, Belgium)

History of the measurement

It is essential for the potential users to understand clearly in which context the instrument has been developed and adopted. A literature review should thus be conducted with a view to answering the following questions:

- What are the most important underlying concepts and what is the public health relevance of the domain investigated?
- How was the measurement considered here performed in the past in either specific or general population health surveys?
- Who proposed the instruments used in the past (give the exact bibliographic references)?
- To what extent have those instruments effectively been used, in which surveys (give some examples of surveys with exact references)?

An inventory of the health surveys performed in Europe is available (https://www.iph.fgov.be/hishes). The questions used in those surveys are also available on the same website together with a translation into English. This inventory provided a good opportunity to evaluate the Health Surveys in the European Union (18;19).

An example of such a detailed description is the Euroreves report “Selection of a coherent set of health indicators” (20); this document is available on the website of the European Commission - DG Sanco:

http://europa.eu.int/comm/health/ph_projects/1998/monitoring/monitoring_project_1998_full_en.htm#3

Other examples can also be found in the manuals of some widely used instruments such as the SF-36 Health Survey module (21), the SCL-90-R symptom checklist (22) or the users’ guide for the General Health Questionnaire (23).

Such an inventory should be complemented by a short review of the validity of each of the instruments used in the past: were the instruments developed successful in collecting valuable data? Did they effectively measure what they intended to measure? What were the results of the test-retest reliability checks? Has a standard been defined (independent source of information such as a register by example) and have the data from health surveys been compared with this standard?

It is important to understand how far the instruments used to measure a particular health concept or phenomenon (such as tobacco consumption, for example) have been progressively replaced by others. If different instruments have been used over time, the rationale behind such an evolution should be explained: is it because the new instruments provide a better measure of the same construct or is it because the concept behind the measurement evolved over time?

Was the measurement always separated in a specific instrument or was it included in a generic instrument such as the Sickness Impact Profile (SIP), the Nottingham Health Profile (NHP) or the Short Form Health Survey (SF-36)?

Review of the instruments

It is important for each EU member state to be able to compare national health indicators with those measured in other countries. Similarly it is crucial for the European Commission to be able to produce common health indicators at the European level.

As a first step it is necessary to prepare a comprehensive review of instruments used to measure the same construct, in the existing or previously conducted health surveys from the different European member states. Such a review is called the “survey of surveys” in the report of the Eurohis project (24).

It must be verified if those instruments are comparable. If yes, what are the common features and how far is it possible to compare the results? If not, what are the main differences and how could these be solved (pre-harmonisation of the instruments; post-harmonisation of the data collected)? An example of such an evaluation process can be found in the report produced by Eurostat in 1998 (25).

The next step is to verify if there are any other instruments currently available besides the ones already used in ongoing or previously conducted health surveys within the EU member states. Can one of these instruments be selected for the measurement that is targeted here? If so, what are the main reasons for selecting one of the instruments available? If not, why is it not possible to select one of the instruments available; why is it necessary to develop a new instrument?

If the instrument selected is different from what is used by most of the European countries, evidence should be provided why a new/other instrument is proposed and how it differs from the already existing instruments.

Description of the instrument

If the instrument proposed is new, a detailed description on how it was designed should be given:

- How was the decision process organised as far as the content of the instrument is concerned?
- How were the concepts integrated in the design of the instrument?
- What are the important issues that have been considered?
- What are the important problems that have been faced in the design of the instrument and how have those been solved?
- How many questions have been included and why?
- What is the potential for comparability between different cultures?

An example of such a description of the construction of a new instrument can be found in the Euroreves report “Selection of a coherent set of health indicators, phase II” (26); this document is available on the website of the European Commission - DG Sanco:

http://europa.eu.int/comm/health/ph_projects/2000/monitoring/monitoring_project_2000_full_en.htm#3

Whether the instrument is new or old, it is necessary in any event to provide the detailed characteristics of the instrument: what is measured and how is this achieved? The same holds true if an existing instrument is modified (i.e. if a list of chronic conditions is extended or to adapt the questions on health consumption to better fit health care delivery characteristics).

Characteristics

All the details that will be useful for the adoption of the instrument in a specific survey have to be provided. The following details are required about the instrument:

- is it advisable to position the instrument in a specific part of the questionnaire (i.e. at the beginning or at the end of the questionnaire, etc …)?
- should the instrument be used in a face-to-face or a self-completed questionnaire?
- are proxy respondents allowed? If so, under which circumstances?
- does the instrument include several questions? If so, can some of these be eliminated (optional questions as opposed to core questions) or is the instrument only valid if all the questions are included?
- in which order should the questions be presented?
- has this instrument been developed for the whole population or is it applicable only to a specific target group (what is the target group and why has such target been selected)?
- must the instrument be focused on a specific period of time: what is the time reference and why has such reference been selected?

There is generally not enough scientific proof concerning such issues as the order and the placing of questions. Recommendations must thus often be based on educated guesses rather than on proven facts. It is however important to document and justify such decisions.

The following details are required for each of the specific questions:

- what are the main features that have to be included in each of the question(s) and that need special attention as far as the wording is concerned?
- what are the answer categories (varying answers categories for each of the items) and/or what is the scoring system (same answers categories for each of the items)?

Are there several forms available for the same instrument, for example a long versus a short form (the short form can be used in a multipurpose survey and the long in a more specific survey)? If this is the case, how has the shorter form been constructed and what are the advantages and the drawbacks of the respective forms? An example of construction of a short version can be found in the GHQ manual (23).

Are there different versions of the same instrument? For example, are there self-administered and telephone interview versions with a different layout?

The IPAQ *(International Physical Activity Questionnaire)* website *(www.ipaq.ki.se)* provides a good example of different layouts for the same instrument[3].

The respective interests and drawbacks of each of the forms should be described in detail.

The instrument should also satisfy other criteria, such as neutral language, lack of ambiguity, simple terminology, etc … when a new instrument is developed /recommended, it is thus necessary to give some explanations on how these requirements have been taken into account.

Even if a neutral language is preferable in general, in practice it may happen that social sensitivity on certain issues could cause a negative reaction of the part of the interviewee. Under these circumstances, if it is suspected that neutral language would provide a 'socially acceptable or desirable response', it would be justified to use some 'emotional discharge' although it could have an influence on the answers.

[3] The IPAQ questionnaire has already been used in health surveys in some European member states but some problems have been reported as far as field data collection as well as analysis of the results are concerned.

Example:

In the pre-test of the Spanish Health and Sexual Habits Survey, the question related to being drunk was asked in two different ways:

- directly (*How many days you consider you have been drunk in the last three months?*)
- and another preceded with and 'emotional discharge' (*Almost everybody has consumed alcoholic drinks in some time of his life. In the last three months, how many days you consider you have been drunk?*).

The experience showed that the second version produced better results (the quality of the results was measured through comparison with different data sources).

Stability

The end result at this stage is the production of the instrument in one source language. It will usually be produced in English). However, despite the predominance of English, the original source language could be any European language (such as Finnish, Swedish, Danish, French, German, Spanish, Italian …). The reason for this is that an expert can hardly develop any valid instrument except in his own mother tongue language.

Whether or not the instrument proposed could effectively be recommended for the health surveys in Europe will depend on several factors. One of the criteria to be taken into account in this area is certainly the potential stability of the instrument over time. It is therefore important to give further information about this:

- What are the potential changes in the future as far as the measurement of the phenomenon under study is concerned?
- Is there any ongoing development work in this area?

Detailed documentation of all those steps is of course imperative.

The information about how to use the instrument in the field, about the kind of indicators to be constructed when analysing the results together with guidelines about how to build those indicators will be described in chapter 6, implementation, of the present document.

4. Quality evaluation of the source instrument

(Hanna Tolonen, National Public Health Institute (KTL), Helsinki, Finland)

This section will discuss the quality evaluation of the instrument in the source language. After the instrument in the source language has gone through the quality evaluation, it will be translated to the target languages. If there is any doubt whether the quality evaluation in source language is sufficient, the quality evaluation should be redone for the instrument after it has been translated to the target languages.

The quality evaluation is a process during which the reliability and the validity of the instrument is assessed. Each step of the evaluation should be documented and made publicly available, i.e. preferably published.

From now on, the instrument is referred as questions.

Critical review of the questions

After the preparation of questions is finished, the team who have prepared the questions should fill in the questions themselves to see if they can answer all the questions without any problems. If the team members do not find any problems with the questions, a number of colleagues/family members/friends are asked to answer the questions and report any difficulties. Usually, the biggest problems with the questions are picked up during this process, which does not cost much and is fast to conduct.

When the questions have passed this critical review by colleagues/family members/friends, the actual pre-testing of the questions can be planned and conducted.

Pre-testing

During the pre-testing, the questions are studied from the respondents' point of view. The clarity, comprehensiveness and acceptability of the questions are tested (27;28):

- *Clarity:* Do respondents understand questions correctly?

- *Comprehensiveness:* Are the words/terms used in the questions known by all the respondents? Are all the response alternatives clear and unequivocal for the respondent? Are all required response alternatives listed? Is the question reasonable and is it really needed? Is the length of recall period feasible?

- *Acceptability:* Are questions ethically and morally approved, i.e. are questions not too sensitive? Do questions affect privacy? Is the respondent's burden acceptable?

The required number of subjects for the pre-testing is somewhere between 25 and 75. Persons selected for the pre-test should have the same background profile as the target population of the survey (27;29;30). The interviewers used in pre-testing should not solely be top professionals but preferably range from professionals to beginners. The top professional interviewers may not reveal all shortcomings of the questions due to their high experience which can compensate some problems related to the questions (30).

Several methods can be used for pre-testing of the questions.

1. Simple testing

The conventional way of conducting a pre-testing of questions is to ask a few experienced interviewers to conduct a small number of interviews using the questions under test and to report their experiences about them (31).

2. Cognitive testing

The cognitive testing is conducted using cognitive interviewing (32;33). In cognitive interviewing the respondent thinks aloud while processing the question and decides how to answer to the question. There are several different ways to conduct cognitive interviewing:

- Respondent verbalizes his/her thoughts while he/she answers to the questions.
- Respondent first gives the answer to the questions and then tells how he/she ended up with the given answer.
- Semi-structured discussion about the questions.
- Respondent tells in his/her own words how he/she understood the question.

3. Behaviour coding

During behaviour coding (34) each interview is monitored and preferably taped for closer evaluation afterwards. From each interview, the interviewer and respondent behaviour to each question is coded. For the interviewer, three different codes are used:

E	Exact	Interviewer reads the question exactly as printed
S	Slight change	Interviewer reads the questions changing a minor word that does not alter question meaning
M	Major change	Interviewer changes the questions such that the meaning is altered. Interviewer does not complete reading the question.

and for the respondent seven different codes are used:

1	Interruption with answer	Respondent interrupts initial question reading with answer.
2	Clarification	Respondent asks for repeat or clarification of question, or makes statement indicating uncertainty about question.
3	Adequate answer	Respondent gives answer that meets question objectives
4	Qualified answer	Respondent gives answer that meets question objectives, but it may indicate uncertainty about accuracy.
5	Inadequate answer	Respondent gives answer that does not meet question objectives

6	Don't know	Respondent gives a "don't know" or equivalent answer
7	Refusal to answer	Respondent refuses to answer the question.

Codes S and M for interviewer and codes 1,2, 4-7 for respondent indicate a potential problem with a question.

Eight different problem indicators can be calculated from the above table (excluding the non-problem codes) by summing up the number of times each code is given for the question.

Example: The question to be studied using behaviour coding is *"Have you been told by a health professional in the past year that you have hypercholesterolemia?".* For the behaviour coding, six interviewers realised 20 interviews each, for total of 120. The overall result for the question was

Problem indicator	Number of times for the questions	%
Interviewer question-reading behaviour:		
S Slight changes	15	12.5
M Major changes	3	2.5
Respondent behaviour:		
1. Interruption	5	4.2
2. Clarification	20	16.7
4. Qualified answer	7	5.8
5. Inadequate answer	25	20.8
6. "Don't know"	13	10.8
7. Refusal	2	1.7

The above results indicate that in 15% of the cases the interviewer changed the wording of the question. This may indicate that the question is difficult to read; in particular the word "hypercholesterolemia" may cause problems for some interviewers.

From respondents, 17% asked for clarification to the question and 21% gave an inadequate answer, indicating that question is too difficult for a large proportion of respondents. It might help to change word "hypercholesterolemia" with the phrase "high total cholesterol".

The behaviour coding cannot identify problems when the respondent gives an acceptable answer to the question but has misinterpreted the actual question or when the respondent chooses to answer without asking for clarification even though they did not understand the question.

The behaviour coding can also be used to compare two different wordings of the same question.

4. Special probing

Special probing (34) can be used only for a few questions at a time or otherwise the length of the interview is increased unduly. The probes can be related to the comprehension of the question, retrieval of the information, selection of response category or generally to the question.

The probes are placed so that they follow the actual question: they can be very general or rather specific, depending on the required information. The use of special probes during the interview is much like a cognitive interview during which more information from the respondent is required.

Example: The question under study is *"Have you been told by a health professional in the past year that you have hypercholesterolemia?".*

- Comprehension probe: “Please, tell me what you understand by health professional, i.e. list all the alternatives”
- Information retrieval probe: “How did you retrieve this information”
- Response category selection probe: “When answering the question, were you able to find the exact response category from those listed or did you have to choose between two or more response categories?”
- General probe: “Could you tell me more about that?”

5. Expert panel

In the expert panel (31) method, the questionnaire is given to at least two independent expert panels for review. Expert panels include people with knowledge of substance of the questions, questionnaire design issues, and the cognitive response behaviours. Each expert panel reviews the questions and lists all possible problems in them. After that, a detailed discussion about the listed problems is held and recorded. Finally, the discussion is summarized in a final report.

6. Comparison of pre-testing methods

Presser & Blair (31) have compared different pre-testing methods and their costs. They found that conventional pre-testing and behaviour coding are the most sensitive methods for identifying possible interviewer problems while cognitive interview hardly detected such problems. Expert panels and behaviour coding were able to detect the widest range of problems with questions.

In the trials of Presser & Blair (31), the expert panel method was cheapest, followed by cognitive interviews. The conventional pre-testing and behaviour coding were the most expensive in their trials. However, the difference in cost between cognitive interviews, behaviour coding, and conventional pre-testing was not significant.

For the cost-effectiveness point of view, the expert panels are best, since they are the cheapest and also detect most of the problems with the questions (31).

Reliability

Reliability of question refers to the consistency of the instrument, i.e. to how well the question produces the same answer from the same respondent when repeated by the same or different interviewer.

To test the reliability of questions, the same questions are asked from the same respondents more than once over a certain period of time by the same interviewer or by different interviewers. Based on data collected by repeated measures of same individuals, the reliability of questions can be assessed using several different methods (35-37).

Please note, that the question can be reliable without being valid. In cases like this, the question produces same answer for repeated administrations in same respondents, i.e. the reliability of the question is high but the question does not measure the actual outcome of interest, i.e. the validity is low.

Validation

The question is valid when it measures the correct outcome, i.e. the answer to the question provides accurate information about the behaviour, phenomena, concepts of health etc. under study.

This is also often referred to as the internal validity of the question, being the extent to which a question, scale or instrument measures the relevant concept, attribute or property. This can be opposed to the external validity that refers to the issue of whether the findings of the survey can be generalised to the whole population. External validity is usually related to the survey methods whereas internal validity is dependent on the question/instrument.

Even if an instrument has shown some problems as far as internal validity is concerned, it could still be used to assess relative differences. I.e. the relations between levels of prevalence between socioeconomic groups could be valid in different countries although the absolute prevalence might be biased.

In the validation process, the data obtained by new questions are checked against other existing data sources. These other data sources may be administrative registers, other available registers, other questions, etc. The basic idea is that the same information, for example use of antihypertensive drug treatment, is retrieved and compared using two data sources, one being new question(s). There are various statistical methods that can be used to assess the validity of the results (37).

For some outcomes, so called “gold standards” exist. These are measures that are used to obtain the true values. To use “gold standards” may be expensive and unfeasible for survey itself as well as for the validation process. However, in an ideal situation, new question(s) are validated against existing “gold standards”.

Where “gold standards” do not exist or are too expensive to be used in the validation process, the validation should be done against other existing data sources. In this case, the selected other data source is considered to provide the true values.

When new question(s) are planned to replace already existing one(s) they have to be “validated” against already existing question(s). The “validation” process consists here in asking simultaneously the “old” and the “new” questions to the same respondents in order to able to compare the outcomes. This is important, since survey organizers won’t adopt new question(s) unless they are sure that their trends won’t be affected by the modification of the question(s).

In testing the validity of a question, the administration mode and interviewer instructions are as important as testing the validity of questions themselves. Sometimes question(s) may require some modification depending on the administration mode. For self-administered question(s), all necessary clarifications have to be written down next to the question itself, while question(s) administered by interview are usually without any additional clarifications and the interviewer has separate instructions on how to probe if the respondent requires help.

Pilot testing (field testing)

The preceding remarks concerning pre-testing, reliability testing and validation have focused on individual questions or instruments. Pilot field-testing, on the other hand, concerns the complete questionnaire to be used in the final survey. Pilot testing is not concerned with how individual questions are understood by respondent and what kind of probes or clarifications for questions is needed. These issues should have been tested and settled during the pre-testing.

During the pilot field-testing, the following items are assessed:

- order and location of questions in the questionnaire
- how well potential jump rules in the complete questionnaire work
- length of complete questionnaire and the time taken to fill it
- respondent burden.

The pilot testing is intended to be a test of overall survey process, covering the measurements (questionnaires and physical measurement), and entire survey logistics and organization (28;38).

For pilot testing, a sample of 100 to 200 should be selected (38). The detailed methodology of the planned survey should be applied and the sample of individuals chosen should resemble the final sample as closely as possible. The data collection method used during the pilot field-testing has to be the same as in the planned survey: face-to-face interview, computer assisted interview or telephone-assisted interview.

It is not possible here to describe in detail either the methodology of pilot testing or the utilisation of the results in the implementation of the population health survey. Those who are planning to execute a pilot field survey are invited to consult the literature available on this subject (39;40). Pilot testing is also described in the chapter on “good survey methodology practices” in the EMCDDA publication (41).In addition, a book on methods of testing and evaluating survey questionnaires (42) has been published recently (July 2004).

Conclusion

Quality evaluation of questions is done to ensure that the results obtained are reliable and valid. It is recognized that pre-testing and validation of the questions are time consuming and expensive processes but nevertheless they form an important part of the question development. To ensure at least the minimum level of reliability and validity of new questions, a pre-testing should be conducted using the most appropriate method for the questions taking into account the financing available.

The reliability and validity of question should be evaluated. If the validity cannot be measured against a “gold standard”, it should be done at least against one other data source. In cases where new questions are planned to replace already existing ones, they should always be validated against the existing questions.

Pilot testing can be left to the survey organizer to test the question(s) in the context of the survey design and the complete questionnaire.

5. Translation

(Christa Scheidt-Nave, Robert Koch Institute, Berlin, Germany)

Introduction

The aim of translation guidelines is to assure the technical, linguistic and conceptual equivalence of health interview questions used in comparative multi-lingual survey research. This 'Ask-the-Same-Question (ASQ)' approach is relevant to international surveys as well as to inter-cultural studies within a given country. The original instrument and all translated versions are expected to 'capture' a particular phenomenon in the specific target populations with consistent reliability and validity. Otherwise the comparability of data collected is not achieved, and the validity of conclusions drawn from the study results will be compromised.

Standardized procedures are needed for the translation process as well as quality assessment of the translated instruments. Translation has to consider linguistic subtleties (e. g. semantic or lexical ambiguity) as well as differences in cultural background in order to avoid misunderstanding. Evaluation procedures need to assess both quality of translation and quality of instrument performance in the translated version and respective target populations.

The aim of this chapter is to provide an overview of existing translation guidelines and protocols developed in previous multi-lingual studies for comparative health assessment. Recommendations for the translation protocol development process will be based on the best available evidence, existing resources will be identified, and controversial / unsettled issues will be discussed.

The end result at this stage will be the translation from a source language to one or more target languages. The source will usually be in English. However, because an expert can hardly develop any valid instrument except in his own mother tongue, the original source language could in fact be any European language.

The present chapter is based on the assumption that a “brand new” instrument has been developed. Nowadays, the general situation is that various established measures exist and possible new ones must build on them.

In fact, when an instrument is proposed, it has frequently already been used in some settings. A preliminary step consists thus in verifying where it has been applied and if any translation procedure has already been performed. Where that is the case, the quality of the translation process should be verified.

Guidelines for Translation Protocol Development

Approach to the translation procedure

There is no ‚gold standard‘ on how to proceed as evaluation studies dealing with the problem of multi-lingual survey research are scarce and different studies have very different needs (43); see also the website of LE et al. http://latino.rcm.upr.edu/spantran.pdf.

The most commonly cited approach to the design of a multi-lingual survey instrument is translation from a source language to one or more target languages (43). In this case, a set of unequivocally phrased questions in the source language has to be already worked out and accepted by all national coordinators represented in the project steering group (43;44). Furthermore, the group of national investigators has to agree on the source language (e.g. English) and a set of target languages. A complete inventory of target languages needs to be taken. The selection of target

languages depends on (a) the official language(s) of participating countries, and (b) the decision to include within-country minority populations who do not speak the first official language(s).

As an alternative to the source-to-target approach, a multi-lingual survey instrument could be developed using an in parallel design (43). For instance, in the annual Eurostat social survey on Statistics on Income and Living Conditions (SILC), current practice is to provide members with a concept definition of each question (and answer), but the actual wording of the questions is left to their own discretion.

Specific recommendations relate to countries that share a main language, such as German for Germany, Switzerland, Austria or English for the U. K. and Ireland.

To conserve resources and render the translation process efficient, a split translation approach and joint translation process referred to as harmonization has been suggested (44;45). This means that countries sharing a first language appoint one translator each to translate part of the instrument. A common protocol for the conduct and documentation of the translation, review and adjudication steps is worked out with the overall aim to agree on one final version that accommodates input from all countries.

With either approach, it is mandatory to have worked out and to agree upon the concepts of health that are to be measured in the survey (see chapter 2).

Coordination of the translation procedure

As documented in the literature, it is highly recommendable to centralize coordination, monitoring and evaluation of the translation process in conjunction with national study coordinators (44), in order to assure adherence to the protocol and rigorous quality control.

Furthermore, it is advisable to consult with and get support material from expert panels who have previously established translation guidelines and protocols for multi-lingual studies. Examples for such resources include the Translation Expert Panel of the European Social Survey (ESS) convened by Janet Harkness at the Centre for Survey Research and Methodology (ZUMA), Mannheim, Germany, and others are available on the following websites:

http://naticent02.uuhost.uk.uu.net/methodology/translation_strategy.htm
http://naticent02.uuhost.uk.uu.net/ess_docs/translation.doc

Another useful example is the Translation Guideline Development Group for the Minimum European Health Module (MEHM); this was organised within the framework of the Euro-REVES 2 project “Setting up of a coherent set of health expectancies for the European Union” (46;47).

After deciding on the scope of the translation process, the details of the translation protocol have to be worked out. Costs need to be calculated and to be explicitly included into the proposed budget.

Main steps of translation procedure

Most experts in the field recommend a team approach including a sequence of steps and a combination of techniques (40;44;48;49):

Step 1	Initial or **forward translation** of questionnaire from source to target language(s) by one or (preferably) at least two independent translators
Step 2	Independent **review** of the initial translation product
Step 3	**Committee/panel adjudication** (deciding on a final version of the translated instrument)
Step 4	**Back translation** (controversial according to current literature)

Forward translation

Who should translate ?

It is recommended to recruit professional, experienced translators from translation associations, colleges or professional network channels (e. g. European Social Survey (ESS) translation work package network channel). Using translation agencies (bureaux) is sometimes not the best choice for various reasons, e. g. higher costs, lack of flexibility and limited interaction (44). Translator applicants should be interviewed and tested for their translation skills. According to expert opinion and experience, translators should fulfil the following requirements (40;43-45;48;49):

- specific ‚target' language (e.g. Austrian-German not just German) as a mother tongue and fluent in 'source' language (English),
- high level of practical translation experience,
- high degree of cultural embedding,
- openness to study concept and team approach, and
- willingness to learn and undergo training specific to the particular requirements of the study.

See also Matias-Carrelo et al on the following website:
http://latino.rcm.upr.edu/spantran.pdf.

An additional health / social science background and previous experience with questionnaire translation may or may not be helpful. On the one hand, translators familiar with health issues may more readily grasp the concepts under study (47). On the other hand, translators with a strong health science background may hold their own, independent view of things, which could result in just the opposite, i.e. considerable deviation from the underlying health concepts (44).

Public health professionals and researchers can sometimes be solicited in the translation procedure instead of independent translators but one must monitor the process carefully and verify they do not intrude their own views and deviate from the original concept.

How many translators?

It is highly recommended to appoint at least two independent translators *(40;43;44;48-51).* Their work will have to be closely guided and monitored; appointed translators need to be carefully briefed about:

- the study goal,
- the background, function and technical components of the original questionnaire (cultural context and time of origin; phenomena/dimensions under study; measurement scales),
- the target population (age range; factors likely to determine response; specific subgroups, e. g. minority groups included), and

- the preferable 'tone' of language (simple vs. sophisticated; casual vs. official).

Ideally, translators should receive further training based on examples, previously used questionnaires etc.

Conceptual translation cards?

Conceptual translation cards need to be provided along with the original study questions in the source language (similar to written instructions for interviewers) to guide the translation process. These instructions will have to be read and translated into the target language b e f o r e translation of the actual instrument; this is to assure that the underlying concept to be measured has been understood (47).

Monitoring the translation process ?

The actual translation work needs to be c o n c u r r e n t l y documented with explicit reference to any particular translation problem (linguistic or conceptual). For this, a *translation documentation template* needs to be designed. This should include rating scales as well as space for free text comments.

If the questionnaire where the module is used will be computerized, translation and documentation have to accommodate another level of complexity, i.e. additional instructions for coding, handling error messages etc.

Translation work should be closely monitored. An interim review is recommended on completion of the first 10 percent of the first assignment (44).

Independent review

Who should be a reviewer (checker) ?

One independent reviewer per target language is considered sufficient. In addition to the same qualifications as the translator(s) (i.e. bilingualism, with target language as first language), reviewers should be familiar with all aspects of the study and questionnaire design (44). If it is impossible to find individuals with expertise in both areas, two different people suitable to fulfil the tasks need to be recruited.

Again the review process needs to be carefully documented in the translation documentation template. Failure to comply would render the reviewing process useless; as a worst case scenario, a second translation may be performed without giving any reasons for the differences (47).

Committee/panel adjudication

Final consent on the translation product should be reached by a panel of experts, preferably including the reviewer, the translator(s) and the adjudicator (usually the national study coordinator). Adjudicators' main field of expertise should be the study objective and design. If they are not proficient in the source language, a consultant should assist them.

Detailed recording of all issues discussed and decisions made at this final step is crucial to keep the translation process transparent and assure the quality of the translation product. Thus, the translation documentation template needs to cover this step appropriately.

Back translation

Back-translation is recommended by some (40;48) but not all experts (44;47;49). Previous experience has demonstrated that formal back translation shifts the focus back on literal translation issues and does not necessarily serve the goal to produce a conceptual equivalent of the original instrument (44;47;49-51).

Field testing and evaluation of translation products

Translated instruments should be tested and evaluated with respect to equivalence in five dimensions: semantic, content, technical, construct, and criterion (see the Matías-Carrelo et al. website: http://latino.rcm.upr.edu/spantran.pdf). The first three relate to translation quality, the latter two to performance quality. Basically, the same guidelines and methods apply as for quality assessment of the source instrument (see chapter 4).

Although time- and cost-consuming, evaluation of the translated instruments in addition to quality assessment of the source instrument is strongly recommended (45;48;49)*;* see also the document by Matías-Carrelo et al. on the following website: http://latino.rcm.upr.edu/spantran.pdf. As has been recently suggested for a translation pilot of three other modules for a European Health Interview Survey, evaluation should include at least the following two steps:

- a person fulfilling all the criteria required for a translator / reviewer, but not familiar with the underlying health concepts should review the final translation product and derive the underlying health concepts from it;
- interviewer and respondent debriefing, with the goal to obtain information how the interviewee understood certain questions and how he/she views the underlying health concepts.

The results of this work are then compared to the concepts that were meant to be measured in the first place. It is well possible that insight gained from the evaluation process will lead to adaptations of the original instrument.

Checklist for translation procedure

§ Is there agreement with respect to the health concepts under study ?
§ Is there agreement with respect to the basic approach towards translation (source-to-language vs. in parallel translation process) ?
§ If a source-to-target approach is chosen - is there agreement on an original instrument including not only the actual health questions but also the text for introductory sentences, filters, and conceptual translation cards ? If the instruments are to be computerized, additional instructions, response categories etc. may have to be translated.
§ Is there exact agreement on the source and target language(s) ?
§ Are there any countries sharing a first main language ? These need to indicate their wish for joint or split translation (harmonization).
§ Who is going to be in charge of coordinating the translation process in cooperation with national study coordinators ?
§ Are there any existing resources that could be consulted for logistic support, i. e. previously developed translation protocols, guidelines or support materials provided by translation task forces or guideline development panels ?
§ Has the instrument already been used in some settings? Has any translation procedure already been performed? What was the quality of the translation process?
§ Is there agreement on the main steps of the translation process, i.e. forward translation / review / adjudication / back translation ?

§ Have documentation templates been designed and agreed upon ?
§ Is there agreement on the evaluation process ?
§ Have a budget and timetable with milestones been prepared and agreed upon to be included in the study proposal ?

Recommendation

It is necessary to make an official deposit (within Eurostat?) of the exact wording of the instrument in the source language, of the conceptual translation cards as well as of the official translation in the different European languages … this is the only way to avoid the circulation and the use of several different versions of the same instrument!! This deposit should be accessible via Internet[4].

[4] The "HIS/HES database" could serve in the future as deposit / library for the reference instruments and the official translations; see : https://www.iph.fgov.be/hishes.

6. Implementing the instrument in the survey and procedures for analysis

(A. Tinto, National Institute for Statistics - ISTAT, Roma, Italy)

Introduction

Implementing a specific instrument in a survey involves various considerable tasks. The potential users should be provided with all the necessary technical information for an effective and correct use of the proposed instrument, including details on:

- data collection method
- training for interviewers
- response rate
- data entry procedures
- data management phases
- data analysis and reporting.

These technical guidelines should take the form of a *User's Manual*. The items to be covered in such a manual are described here in detail.

The User's Manual should include information related to three main phases: the data collection on the field, the data entry and management, and the data analysis and reporting.

Field data collection

The choice of how to administer the questionnaire needs to be considered carefully, as different **data collection methods** may produce different results, and the instrument used in a self-completed questionnaire may not be completely identical to the one that should be used in a face-to-face interview, for instance.

The different techniques that can be used in a Health Survey are, in particular: *self-administered* (by post, internet or with the presence of a researcher), *face-to-face* (Computer Assisted Personal Interview-CAPI or Paper and Pencil Interview-PAPI), and *telephone interview* (of which the most common type is the Computer Assisted Telephone Interview - *CATI*).

The list of advantages and disadvantages for each technique presented in Table 6.1 may be useful to assist in the choice of the most appropriate data collection method (52).

If the instrument to be implemented requires strict rules about the data collection method to be used, these should be described in detail in the User's Manual together with the reasons why these rules should be followed carefully. As an example, the SCL-90-R symptom checklist (22) recommendations indicate that it can only be administered through a self-completed questionnaire (paper and pencil or on-line).

Other modules may have less strict requirements regarding the type of data collection method. In this case, or when the survey is composed by different parts using different techniques, indications should be given in the User's Manual about the best way to administer the questions. This will depend on:

- the nature and the purpose of the study;
- specific recommendations related to the instrument (for example if the questions include predominantly delicate matters, then it would be more advisable to use a self-completed questionnaire);

- characteristics and needs of the respondents: it is crucial that the participants to the survey remain interested and cooperative throughout the whole interview, understand the questions in an univocal way and that they receive the appropriate support during the completion of the questionnaire;
- resources available (both in terms of funds and skills);
- time constraints.

Table 6.1

TECHNIQUE	ADVANTAGES	DISADVANTAGES
Self-administered		
by post	Lower costs Fast to reach large samples Lower non-sampling error due to the interviewer effect	Misunderstandings cannot be checked Complicated issues cannot be clarified Limited use of filters and skip rules Relies on the mailing system Relies on the literacy skills of the respondent Lower response rate
by email/website	Easy to design and send out Keeps track of non respondents (reminders can be sent)	Same as post plus: Sample bias (only those with internet access can be sampled)
Self-administered with interviewer present	Interviewer is there to offer help when needed Interviewer can check	The presence of the interviewer could 'lead' participants Questionnaire needs to be self-explanatory to minimise discretionary interpretation
Face-to-face		
PAPI	Misunderstandings can be discussed Does not require literacy of respondent Produces less missing data Possible to detect information from direct observation	Higher costs Interviewer effect Difficult to ask sensitive questions Time consuming for large samples
CAPI	Same as PAPI plus Possibility of automatic interview paths, checks and data entry	Same as PAPI
By telephone (CATI)	Lower costs Fast to reach large samples Automatic interview paths, checks and data entry Personal (sensitive questions can be asked)	Sample bias (only those in telephone directory can be sampled) Lower response rate Higher non-response for selected groups of population Increasing number of mobile phones are changing the nature of telephone interviews to personal instead of household surveys and creating problems for sampling frames.

The User's Manual should also contain detailed explanations on the consequences of each choice and on the steps that need to be taken in order to give greater value to the positive aspects linked with the selected data collection method, and minimize the disadvantages connected to it (53). In particular, these are the areas that need to be considered:

- layout of the questionnaire
- instructions for completion of the questionnaire
- training of the interviewers.

Layout of the questionnaire

Any specific layout characteristics related to the instrument proposed, that could be useful to increase the quality of the data collection, should be clearly described in the User's Manual. In general, if there are no specific indications, the decisions on the layout should follow the rules specific to the data collection method used in the questionnaire.

Instructions for completion

Specific instructions for completion related to the instrument proposed, that could be useful to increase the quality of the data collection, should be clearly described in the User's Manual. For example, if the instrument to be implemented is *self-administered*, indications should be given on the comments and/or instructions that need to be reported next to each question. To increase the quality of the data, in case of *face to face* and *CATI* data collection method, care should be exerted on the instructions for the interviewers.

Training for interviewers

When there are specific instructions related to the instrument proposed to be provided to the interviewers, another phase immediately preceding the field data collection, which needs to be accurately described in the User's Manual, is the **training for interviewers**. This phase is extremely important to maximise the quality of the collected information, by communicating the survey's contents and goals (definitions, classifications, research aims) as well as difficult situations specific to the instrument that may occur.

By exerting care in these phases of the research, a high level of data quality can be obtained. The best indicator to measure how much confidence can be placed in the results for the specific instrument is the **response rate.** The potential for bias increases as the response rate decreases. Problems in the comparison between countries / surveys exist, as often there is no uniformly applied definition of response rate.

Groves and Couper provided a conceptual framework that classifies factors influencing the response rate by the degree to which they can be controlled by the researchers. Factors that researchers cannot control are the *social environment* or the characteristics of the sample *households*. Factors that the researcher can influence are the *survey design* and the *interviewer characteristics (54)*. The controllable and uncontrollable factors combine to determine the quality of the household-interviewer interaction, which in turn influences the response rate. Although the environmental and household factors are beyond the control of the researcher, an awareness of their role in response rates can help in the proper survey design and interviewer recruitment (55).

Koponen and Aromaa (56) also reviewed the factors supporting or reducing response and participation in health surveys. They mention three factors found to be relevant to the recruitment of diverse populations into research studies:

- **Awareness** defined as an understanding of the importance of research, the procedures during the research process, and the value of the individual's participation (57).
- **Acceptability** defined as social support for participation, reflected in the messages disseminated by community leaders and through media
- Promoting **access** means reducing the practical barriers to participation, e.g. through transportation, understandable consent forms, translation services in multilingual populations, and financial remuneration.

In the User's Manual for a specific instrument it is therefore important to give an indication of the 'desirable' level of the response rate, based on the rate obtained in other surveys.

As non-respondents are not, in general, a random sample, it is essential to include in the User's Manual for the instrument proposed explicit advice on how to maximise the response rate and how to conduct non-response analyses in order to be able to correctly interpret the results.

Quantifying refusals and understanding the reasons behind them is very important, as different strategies for interventions correspond to different types of refusal.

Factors known to increase response rate for a specific instrument are, for example, adopting a simple layout and clear design of the questionnaire and, choosing accurately the data collection period.

Data entry and data management

As entering, checking and cleaning the data is an extremely delicate phase, it is important that details are given to the users on the main steps to follow in order to have the data as 'clean' as possible.

Data entry

The system for performing data checking, cleaning, and recoding should be planned from the beginning, and it should be universally agreed and understood by all researchers involved in the study and be well documented. In the **data entry phase** all individual answers from the questionnaire should be entered in the data file according to the values that respondents or interviewers recorded in the questionnaire. It could happen, though, that the value to be registered is not so easily identifiable. It is thus important to supply in the User's Manual any useful specifications on how to deal with these situations. In the case of the SF-36 Health survey module, for instance, a set of rules were identified (21):

- if the respondent chose two adjacent response categories, the value of one of the two, randomly selected, should be recorded.
- if the respondent chose two nonadjacent response categories for the same question, a missing value should be recorded.
- if the respondent chose three or more response categories for the same question, a missing value should be recorded.
- If the respondent wrote the answer instead of circling the chosen answer category, then the value should be recorded as if the response had been circled.

Recoding

An important phase that follows data entry is the **recoding of the responses**, which allows the user to have the values needed for the calculation of the scores. The first step of this process is to verify whether there is any question with **out-of-range response values**. These are values lower or higher than the minimum and maximum threshold value determined for that specific question. If the problem is not due to a data entry error, it is necessary to give advice on how to deal with it, for example give instructions that these out-of-range values should be changed into missing values. Such instructions should be provided in the User's Manual and each change done to the original data should be well documented.

If the instrument involves a **scoring system** to build synthetic indicators, it is important to give extremely clear and detailed instruction on how to calculate those scores, and on the characteristics and use of the different indicators that, in certain cases, can be produced.

Examples on this can be found in the manual of the SF-36 Health Survey module (21), or the SCL-90-R symptom checklist (22) or the users guide for the General Health Questionnaire (23).

If appropriate, the issue of licence/copyright for the scoring and/or analysis procedures should be also mentioned in the User's Manual.

Item non-response

Item non-responses occur when only part of the information related to a specific unit is missing. This happens, for instance, when the interviewer forgets to ask or record the answer to a question or when the interviewee is not able or does not want to answer to a specific question. This kind of

non-response has to be carefully considered. In particular it is important to understand the possible underlying mechanism determining the item non-response, which can be related to the questions of interest.

If any question in the proposed instrument is more likely to induce such a problem, this should be mentioned in the User's Manual, together with potential solutions to possibly reduce the specific item non-response rate.

There are various methods which can be used to reduce the bias of the final survey estimates resulting from items non-response. Advice should be given on the best technique to be used for the specific instrument or on the advantages and disadvantages related to the different methods that are available. It is particularly important to give all the detailed instructions on this matter especially when a score has to be built.

A method commonly used is **imputation**. It consists in assigning substitute values to the missing data, to be able to restore the complete data matrix (58;59). Some researchers (60) believe that the multivariate nature of statistical surveys, in which any variable could potentially present a missing value, justify the use of imputation in order to reduce the bias due to item non-response in the estimates and to have a complete data set. In any case, the possible disadvantages of imputation need to be considered, and it is necessary to provide detailed instructions on when imputation is suggested, on the method to be used together with the best software to be used to apply this method.

The literature suggests several imputation methods for items non-response, which can generally be grouped into three categories:

- *deductive methods*, the imputed value is deduced from known information or relations;
- *deterministic methods*, repeated imputations for units with the same characteristics always produce the same imputed values;
- *stochastic methods*, repeated imputations for units with the same characteristics could produce different imputed values; they are characterised by the presence of an aleatory component (residual), corresponding to a probabilistic scheme associated with the chosen imputation method.

The impact of the item non-responses on the building of scores and indices should also be mentioned.

Data analysis and reporting

The set of questions included in a specific instrument is usually aimed at the construction of a specific set of **indicators or indices** rather than at the analysis of each of the single questions. If this is the case, detailed information should be provided in the User's Manual on how to use the collected data and produce the indicators. Something should also be mentioned about the treatment of outliers.

For instance, height and weight are measured in the population to build the Body Mass Index, a composite index commonly used to measure body composition. For the sake of international comparability, the formula to calculate the BMI (Kg/m^2) has to be clearly explained, together with the cut-off values used to classify the population as obese or overweight should be provided, and appropriately referenced. Instructions on how to deal with outliers are also advisable.

Before anyone starts working on the dataset, the **statistical methods** to be used for the analysis should also be agreed on. Even if several packages usually yield similar results, it is important to verify if they make it possible to take adequate account of the survey / sample design and the

weighting factors in order to produce correct estimates and variances. Here is a non exhaustive list of such packages: SAS, SUDAAN, StatA, R and Spss (V12 and higher).

In addition, when joint / European publications are considered, details should be given on how data should be analysed and presented

- which are the basic tables to be reported
- how the variables or indicators built for the specific instrument should be categorised
- which background variables such as age, gender, education income, …) should be taken into account.

All those points are important for obtaining comparability.

Conclusions

In conclusion, the main phases needed for a correct implementation of the proposed instrument, can be described in the following list, which could be considered as the table of contents of the User's Manual.

1. Data collection in the field
   1.1 Data collection method
   Data collection method to be used (self-administered, face-to-face, and telephone interview).
   Reasons why the choice should be made.
   1.2 Layout of the questionnaire
   Specific layout characteristics related to the instrument proposed.
   1.3 Instructions for completion of the questionnaire
   Instructions for the interviewers to be included in the questionnaire or instructions for the respondents to be included in a self-completed questionnaire.
   1.4 Training of the interviewers
   Specific instructions related to the instrument to be provided to the interviewers.
   1.5 Response rate
   'Desirable' level of the response rate for the instrument.
   Advice on how to maximise the response rate and how to conduct non-response analyses in order to be able to correctly interpret the results.

2. Data entry and management
   2.1 Data entry procedures
   How to deal with cases when the value to be registered is not easily identifiable.
   2.2 Recoding
   Treatment of out of range values and recoding needed for the calculation of scores.
   2.3 Scoring system
   Calculation of scores.
   Description of the characteristics and use of the different indicators that can be produced.
   2.4 Item non-response
   Potential solutions to reduce the specific item non-response rate.
   Impact of the item non-responses on the building of scores.
   Details of the possible use of imputation techniques.

3. Data analysis and reporting
   3.1 Data analysis and reporting
   How to build indicators from the collected data.
   Preferred statistical methods to be used.

7. Reference list

(1) Rose GA, Blackburn H, Gillum RF, Prineas RJ. Cardiovascular survey methods. 2nd ed. Geneva: WHO, 1982.

(2) Aromaa A. Health observation and health reporting in Europe. Rev Epidemiol Sante Publique 1998; 46(6):481-490.

(3) Van Oyen H, Tafforeau J. Health Interview Survey. Arch Public Health 1994; 52:79-82.

(4) Van Oyen H, Tafforeau J, Hermans H, Quataert P, Schiettecatte E, Lebrun L et al. The Belgian Health Interview Survey. Arch Public Health 1997; 55:1-13.

(5) Kramers PG. The ECHI project: health indicators for the European Community. Eur J Publ Health 2003; 13(3 Suppl):101-106.

(6) Bergner M. Measurement of Health Status. Medical care 1985; 23(5):796-804.

(7) De Kleijn-De Vrankrijker MW. Internationale aspecten van gezondheidsmeting. In: Gunning-Schepers LJ, Mootz M, editors. Gezondheidsmeting. Houten: Van Loghum, 1992.

(8) Mootz M, Van den Berg J. [Indicators of health status in the CBS-Health Interview Survey]. Mndber Gezondheid (CBS) 1989; 2:4-10.

(9) Guide to Clinical Preventive Services: Report of the US Preventive Services Task Torce. 2 ed. Baltimore: William & Wilkins, 1996.

(10) Schaapveld K, Bergsma EW, Van Ginneken JKS, Van De Water HPA. Setting priorities in prevention. Leiden: TNO Institute for Preventive Health Care, 1990.

(11) Robine JM, Romieu I. Setting up a coherent set for health indicators for the European Union/Euro-Reves project for the Community Action on Health Monitoring. 1-8. 1998. Montpellier, INSERM.

(12) Pelletier L. Les indicateurs de santé et la planification sanitaire. Revue Epid et Santé Publ 1990; 38:47-56.

(13) Recht P. Le concept d'indicateur de santé. Actualité et perspectives. Arch Belg 1982; 40(5-6):197-216.

(14) Lewis A, Modle W. Health indicators: what are they? An approach to efficacy in health care. Health Trends 1982; 14:3-8.

(15) Escobar W. An introduction to indicators. World Hosp 1985; 21(1):32-34.

(16) Mootz M. Health indicators. Social Science and Medicine 1986; 22(2):255-263.

(17) Hansluwka HE. Measuring the health of populations. Indicators and interpretations. Soc Sci Med 1985; 20:1207-1224.

(18) Aromaa A, Koponen P, Tafforeau J, Vermeire C. Evaluation of Health Interview Surveys and Health Examination Surveys in the European Union. Eur J Publ Health 2003; 13(3 Suppl):67-72.

(19) Aromaa A, Koponen P, Vermeire C, Tafforeau J. Status and future of Health Interview in the European Union, final report of phase 2 of the project "Health surveys in the European

Union : HIS and HIS/HES evaluation and models". Helsinki: National Public Health Institute (KTL), 2003.

(20) Robine JM, Jagger C, Egidi V. Selection of a coherent set of health indicators. Montpellier: Euro-Reves, 2000.

(21) Ware Jr JE. SF-36 Health Survey: manual and interpretation guide. Boston: The Health Institute, New England Medical Center, 1993.

(22) Derogatis LR. SCL-90-R administration, scoring and procedures manual. Minneapolis: National Computer System, 1994.

(23) Goldberg D, Williams P. A user's guide to the General Health Questionnaire. Berkshire: NFER-NELSON, 1988.

(24) EUROHIS: Developing common instruments for health surveys. Amsterdam: World Health Organisation Regional office for Europe, 2003.

(25) Hupkens C. Coverage of health topics by surveys in the european union. Luxembourg: Eurostat, 1997.

(26) Robine JM, Jagger C, Romieu I. Selection of a Coherent Set of Health Indicators for the European Union, Phase II. Montpellier: Euro-Reves, 2002.

(27) Rea LM, Parker RA. Designing and conducting survey research. A comprehensive guide. 2nd ed. San Fransisco: Jossey-Bass, 1997.

(28) Biemer PP, Lyberg LE. Introduction to survey quality. New Jersey: John Wiley & Sons, 2003.

(29) Sudman S, Bradburn NM. Asking questions. A practical guide to questionnaire design. San Fransisco: Jossey-Bass, 1982.

(30) Converse JM, Presser S. Survey questions. Handcrafting the standardised questionnaire. Newbury Park: Sage Publications, 1986.

(31) Presser S, Blair J. Survey pretesting: do different methods produce different results. Sociological Methodology 1994; 24:73-104.

(32) Tourangeau R, Rips LJ, Rasinski K. The psychology of survey response. New York: Cambridge University Press, 2000.

(33) Smith TW. Developing comparable questions in cross-national surveys. In: Harkness JA, Van de Vijver FJR, Mohler PP, editors. Cross-cultural survey methods. New Jersey: John Wiley & Sons, 2003.

(34) Oksenberg L, Cannell C, Kalton G. New strategies for pretesting survey questions. Journal of Official Statistics 1991; 7:349-365.

(35) White SA, van den Broek NR. Methods for assessing reliability and validity for a measurement tool: a case study and critique using the WHO haemoglobin colour scale. Stat Med 2004; 23(10):1603-1619.

(36) De Vellis RF. Scale development. Theory and applications. 2nd ed. Thousand Oaks: Sage Publications, 2003.

(37) Carmines EG, Zeller RA. Reliability and validity assessment. Beverly Hills: Sage Publications, 1979.

(38) Dilman DA. Mail and Internet surveys. The taylored design method. 2nd ed. New York: John Wiley & Sons, 1999.

(39) The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. Soc Sci Med 1998; 46:1569-1585.

(40) Bullinger M, Alonso J, Apolone G, Leplege A, Sullivan M, Wood-Dauphinee S et al. Translating health status questionnaires and evaluating their quality: The IQOLA project approach. J Clin Epidemiol 1998; 51(11):913-923.

(41) Bless R, Korf D, Riper H, Diemel S, Vicente J, Hartnoll R et al. Handbook for surveys on drug use among the general population. 1-151. 2002. Lisbon, European Monitoring Centre for Drugs and Drug Addiction.

(42) Presser S, Rothgeb JM, Couper MP, Lessler JT, Martin E, Martin J et al. Methods for Testing and Evaluating Survey Questionnaires. Paperback, 2004.

(43) Bhopal R, Vettini A, Hunt S, Wiebe S, Hanna L, Amos A. Review of prevalence data in, and evaluation of methods for cross cultural adaptation of, UK surveys on tobacco and alcohol in ethnic minority groups. BMJ 2004; 328(7431):76.

(44) Harkness J. An outline of the European Social Survey translation strategies and procedures. 2002. ESS documents.

(45) Trevino FM. Cross-cultural aspects. Plan and operation of the hispanic Health and Nutrition Examination survey 1982-84. US Department of Health and Human Services, 1985.

(46) Robine JM, Jagger C. Creating a coherent set of indicators to monitor health across Europe: the Euro-REVES 2 project. Eur J Publ Health 2003; 13(3 Suppl):6-14.

(47) Robine JM, Jagger C. Report to Eurostat on European Health Status module. 200135100021, 1-123. 31-10-2003. Montpellier, Euro-REVES.

(48) Cross-cultural adaptation of survey instruments: the CAHPS experience.: 7th Conference on Health Survey Methods, CDC, NCHS, 1999.

(49) Harkness J, Schoua-Glusberg A. Questionnaires in translation. In: Harkness J, editor. Cross cultural equivalence. Manheim: ZUMA-Nachrichten Spezial Band 3, 1998: 87-128.

(50) Harkness J. Questionnaire translation. Cross-cultural survey methods. New Jersey: John Wiley and Sons, 2003.

(51) Harkness J, Pennell BE, Schoua-Glusberg A. Survey questionnaires translation and assesssment. In: Presser S, Rothgeb J, Couper M, Lessler J, Martin E, Singer E, editors. Questionnaire development evaluation and testing methods. New Jersey: John Wiley and Sons, 2004.

(52) Prattala R, Helasoja V, Laaksonen M, Laatikainen T, Nicander P, Puska P. Cindi health monitor, proposal for practical guidelines. Helsinki: National Public Health Institute, 2001.

(53) Sabbadini,L.L. Telephone surveys. New problems and new solutions in monitoring data quality: the italian experience. Sorrento, Italy: International seminar on new techniques and technologies for statistics, 1998.

(54) Groves RM, Couper MP. Nonresponse in household interview survey. New York: Wiley, 1998.

(55) National Household Survey on Drug Abuse: 1999. Nonresponse analysis report. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Statistics, 2004.

(56) Koponen P, Aromaa A. Survey design and methodology in National Health Interview and Health Examination Surveys: review of literature, European survey experiences and recommendations. 1-109. 2002. Helsinki, National Public health Institute, KTL.

(57) Brown BA, Long HL, Gould H, Weitz T, Milliken N. A conceptual model for the recruitment of diverse women into research studies. J Womens Health Gend Based Med 2000; 9(6):625-632.

(58) Grande E, Luzi O. Metodologie per l'imputazione delle mancate risposte parziali: un'analisi critica e soluzioni disponibili in Istat. 6. 2003. ISTAT. Quaderdi di Ricerca.

(59) New developments in edit and imputation oracticæs - needs and research.: Conference of the International Association of Survey Statisticians, 2001.

(60) Imputing for missing survy response.: American Statistical Association, 1982.