Hanna Tolonen

Towards the High Quality of Population Health Surveys

Standardization and Quality Control

Publications of the National Public Health Institute A 27/2005

National Public Health Institute Helsinki, Finland
and
Faculty of Medicine, University of Kuopio, Finland

Hanna Tolonen

TOWARDS THE HIGH QUALITY OF POPULATION HEALTH SURVEYS

STANDARDIZATION AND QUALITY CONTROL

ACADEMIC DISSERTATION

To be presented with the permission of the Faculty of Medicine, University of Kuopio, for public examination in Auditorium L1, University Building Canthia, on 28 February 2006, at 12 o'clock.

National Public Health Institute, Helsinki, Finland

and

Faculty of Medicine, University of Kuopio, Finland

Helsinki 2006

Publications of the National Public Health Institute KTL Axx / 2006

Julkaisija-Utgivare-Publisher

Kansanterveyslaitos (KTL)
Mannerheimintie 166
00300 Helsinki
Puh. vaihde (09) 474 41, telefax (09) 4744 8408

Folkhälsoinstitutet
Mannerheimvägen 166
00300 Helsingfors
Tel. växel (09) 474 41, telefax (09) 4744 8408

National Public Health Institute
Mannerheimintie 166
FIN-00300 Helsinki, Finland
Telephone +358 9 474 41, telefax +358 9 4744 8408

ISBN 951-740-546-4
ISSN 0359-3584
ISBN 951-740-547-2 (pdf)
ISSN 1458-6290 (pdf)

Kannen kuva - cover graphic: Tiina Laatikainen & Hanna Tolonen

Edita Oy
Helsinki 2006

Supervised by

Prof. Aulikki Nissinen
Department of Epidemiology and Health Promotion
National Public Health Institute (KTL)
Helsinki, Finland

PhD Kari Kuulasmaa
Department of Epidemiology and Health Promotion
National Public Health Institute (KTL)
Helsinki, Finland

Prof. Pekka Puska
National Public Health Institute (KTL)
Helsinki, Finland

Reviewed by

Prof. Seppo Sarna
Department of Public Health
University of Helsinki
Helsinki, Finland

Docent Hannu Vanhanen
Finnish Heart Association
Helsinki, Finland

Opponent

Docent Juha Mustonen
North Karelia Central Hospital
Joensuu, Finland

If you can't explain it simply,
you do not understand it well enough

- *Albert Einstein*

-

Hanna Tolonen, Towards the high quality of population health surveys. Standardization and quality control.
Publications of the National Public Health Insitute, A27/2005, 181 Pages
ISBN 951-740-546-4; 951-740-547-2 (pdf-version)
ISSN 0359-3584; 1458-6290 (pdf-version)
http://www.ktl.fi/portal/4043

Abstract

Population health surveys are conducted to obtain information about the health status and behaviours of the population. This information is needed for the planning of health policies and prevention activities and for their evaluation.

As the examination of the entire population is generally impossible, the population health surveys are usually sample surveys. Selecting only a portion of the entire population for the examination will set high requirements for the survey, since the inference to the entire population is made from the results of this sample. The sample has to be representative for the entire population and the results obtained from the sample have to be reliable and accurate.

The sample surveys have a number of components, which easily can reduce the reliability and the accuracy of the results. The process of the population health survey should be planned and conducted carefully to avoid potential bias of the results. Starting with sample selection, an up-to-date sampling frame with good coverage of the target population is the first step towards the reliable and accurate population level inference. Next, the selection, development, and validation of the survey instruments play an important role in the successful survey process just like the selection and training of the personnel. The selected survey instruments will provide accurate results, when the measurement protocols are well standardized and personnel know how the protocol works. On top of that, the work of the personnel in the field

requires monitoring, through quality control, to ensure that the standardized measurement protocols are followed.

When the health survey is conducted following the standardized protocol and the proper quality control measurers are applied, the reliability and the accuracy of the results are high. However, even with these conditions met, the reliability and the accuracy of the results can still be compromised if the response rate remains low. Low response rates are a serious problem for the surveys, since they not only reduce the accuracy of the estimates but also have an effect on the representativeness of the sample.

The information collected by population health surveys can often not be obtained through existing administrative statistics. In order to have reliable, accurate and representative information about the population health from the surveys, the attention has to be paid to the quality of the surveys. Standardized survey methods and a properly organized quality control system will help to increase the quality of health surveys.

Keywords: Health survey, standardization, quality control, bias

Hanna Tolonen, Tavoitteena laadukas väestön terveystutkimus. Menetelmien yhtenäistäminen ja laadun valvonta.
Kansanterveyslaitoksen julkaisuja, A27/2005, 181 sivua
ISBN 951-740-546-4; 951-740-547-2 (pdf-versio)
ISSN 0359-3584; 1458-6290 (pdf-versio)
http://www.ktl.fi/portal/4043

Tiivistelmä

Väestön terveydentilan ja terveyskäyttäytymisen selvittämiseksi tehdään väestötason terveystutkimuksia. Tietoa väestön terveydentilasta ja terveyskäyttäytymisestä tarvitaan terveyspoliittisten päätösten avuksi ja erilaisten sairauksien ehkäisyohjelmien suunnitteluun ja tulosten seurantaan.

Koska koko väestön tutkiminen on yleensä ottaen mahdotonta, terveystutkimukset ovatkin yleensä otanta tutkimuksia. Otanta tutkimuksilla, joissa vain osa väestöstä valitaan tutkittavaksi, on korkeat vaatimukset, koska johtopäätökset koskien koko väestön terveyttä tehdään pieneen otokseen perustuen. Otoksen on oltava koko väestöä edustava ja otoksesta saatujen tulosten tulee olla luotettavia ja oikeita.

Otanta tutkimukseen liittyy joukko tekijöitä, jotka voivat helposti laskea tulosten luotettavuutta ja oikeellisuutta. Väestön terveystutkimus tulisi suunnitella ja toteuttaa harkiten, jotta mahdolliset harhan lähteet voidaan välttää. Otoksen valinnassa on tärkeää, että otoskehikko kattaa koko kohde väestön. Tämä on ensimmäinen askel kohti luotettavia ja tarkkoja väestötason tuloksia. Seuraavaksi tulee kiinnittää huomio tutkimuksessa käytettävien menetelmien ja tiedonkeruu välineiden valintaa, kehittämiseen ja testaamiseen. Myös tutkimushenkilökunnan valinta ja koulutus ovat keskeisessä asemassa. Valitut tiedonkeruu menetelmät tuottavat oikeita tuloksia, kun menetelmien ohjeet on hyvin standardoitu ja henkilökunta tietää miten menetelmiä tulee noudattaa. Kaiken tämän lisäksi henkilökunnan

työtä kentällä tulee valvoa, laadun valvontaa apuna käyttäen, jotta voidaan varmistaa että laadittuja standardeja ja ohjeita todellakin noudatetaan.

Terveystutkimuksessa, joka toteutetaan noudattaen standardoituja ohjeita ja riittävästä laadun valvonnasta pidetään huolta, tulosten luotettavuus ja oikeellisuus on korkea. Kuitenkin vaikka menetelmien ohjeita noudatettaisiin ja laadun valvonta olisi hoidettu voi alhaiseksi jäävä vastausaktiivisuus kyseenalaistaan tulosten luotettavuuden ja oikeellisuuden. Matala vastausaktiivisuus on vakava ongelma otanta tutkimuksissa, sillä se ei pelkästään vaikuta tulosten oikeellisuuteen vaan laskee myös otoksen edustavuutta väestöön nähden.

Hallinnolliset rekisterit eivät yleensä pysty tuottamaan tietoa, jota voidaan kerätä otokseen pohjautuvilla terveystutkimuksilla. Jotta terveystutkimusten tuottama tieto väestöstä olisi luotettavaa, oikeaa ja koko väestöä edustavaa on tutkimuksen laatuun panostettava. Standardoidut menetelmät ja kunnolla järjestetty laadun valvonta auttavat terveystutkimusten laadun parantamisessa.

Avainsanat: Terveystutkimus, menetelmien yhtenäistäminen, laadun valvonta, harha

Table of contents

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on five original publications, which are referred to in the text using the Roman numerals I to V.

I. **Tolonen H**, Dobson A, Kulathinal S for the WHO MONICA Project. Effect on trend estimates of the difference between survey respondents and non-respondents: Results from 27 populations in the WHO MONICA Project. *European Journal of Epidemiology,* 2005;**20:**887-898

II. **Tolonen H**, Dobson A, Kulathinal S for the WHO MONICA Project. Assessing the quality of risk factor survey data: lessons from the WHO MONICA Project. *European Journal of Cardiovascular Disease Prevention and Rehabilitation*, 2005; in press

III. **Tolonen H**, Ferrario M, Kuulasmaa K for the WHO MONICA Project. Standardization of total cholesterol measurement in the population surveys - pre-analytical source of variation and their effect on the prevalence of hypercholesterlaemia. *European Journal of Cardiovascular Disease Prevention and Rehabilitation*, 2005;**12:**257-267

IV. **Tolonen H**, Ferrario M, Minoja M for the WHO MONICA Project. Quality assessment of data on awareness and treatment of high cholesterol in the WHO MONICA Project. (June 1999). Available from URL:http://www.ktl.fi/publications/monica/hich/hchdrug.htm, URN:NBN:fi-fe199930

V. **Tolonen H**, Keil U, Ferrario M, Evans A for the WHO MONICA Project. Prevalence, awareness and treatment of hypercholesterolaemia in 32 populations: Results from the WHO MONICA Project. *International Journal of Epidemiology*, 2005;**34:**181-92

ABBREVIATIONS

CAPI	Computer Aided Personal Interview
CATI	Computer Aided Telephone Interview
CVD	Cardiovascular Disease
EHRM	European Health Risk Monitoring Project
EU	European Union
GP	General Practitioner
HES	Health Examination Survey
HIS	Health Interview Survey
MDC	MONICA Data Centre
MONICA	Multinational MONitoring of trends and determinants in CArdiovascular disease
NHANES	National Health and Nutrition Examination Survey
PAF	Postcode Address File
RNIB	Royal National Institute for the Blind
WHO	World Health Organization

ACKNOWLEDGEMENT

This study was carried out in the Department of Epidemiology and Health Promotion (ETEO) at the National Public Health Institute (KTL) of Finland. I wish to thank the KTL, and especially the International Cardiovascular Disease Epidemiology Unit (previously MONICA Data Centre) for giving me the opportunity to work on this study.

My special thanks go to my supervisors PhD Kari Kuulasmaa, Prof. Aulikki Nissinen, and Prof. Pekka Puska. Their endless support, guidance, and trust in me during these years helped me to complete this study.

I would like to express my thanks to all my co-authors, who devoted their time to work on the papers with me. Also, my thanks go to all MONICA investigators, who have spent years in collecting the data and made it possible for me to use the data in this study. Special thanks go to Hermann Wolf, Annette Dobson, Hugh Tunstall-Pedoe, Alun Evans, Kjell Asplund, and Vivan Lundberg who have supported me in this work throughout the years. Their warm words of encouragement, tips and hints on small and big issues have been invaluable.

I thank Prof. Hermann Wolf for the revision of English language and for his valuable comments. I also thank the staff of KTL library, for their help in tracking down the background material for this work.

I thank Prof. Seppo Sarna and Docent Hannu Vanhanen for their valuable work as reviewers of my thesis.

I would also like to thank my colleagues at ETEO who have been very supportive and willing to help me with my questions on medical and epidemiological issues and with the technical problems I have encountered.

And finally, my loving acknowledgements go to my parents Salli and Kari, who have never stopped trusting me and have always supported me in my dreams and most of all to Timo, who has stood by me both on good days and on bad days. Timo, your love has carried me though the hard times and has given me the strength to finalize this work.

Vantaa, December 2005

Hanna Tolonen

1 INTRODUCTION

Epidemiology, defined as “the study of the distribution and determinants of health-related states or problems” [1], is a wide research area covering non-communicable diseases, infectious diseases as well as health behaviours and environmental factors of health, just to list few. A commonly used tool in epidemiological research is the population survey.

Ancient Romans and Egyptians already conducted surveys. In the Roman and Egyptian empires, surveys were used as bases for setting tax rates, military conscription, and making other administrative decision. Since the 1700s, surveys have been used to study social problems. The first social surveys focused on the poor and outcasts of society. [2, 3] The history of health surveys is fairly young. The real development and rapid increase of epidemiological studies got started just after the Second World War. [4]

This study will focus on a small fraction of epidemiological research, the determinants of the cardiovascular diseases (CVD), but many of the results are applicable to the wider area of health surveys. Cardiovascular health surveys provide a useful tool for the collection of information about the determinants of cardiovascular diseases, and their development at the population level. Often this kind of information cannot be obtained through existing official or administrative data systems, like hospital discharge registers or mortality statistics.

There are two main types of health surveys; either collecting information by questionnaire or by examination. In the questionnaire-based surveys, the information is collected using only questionnaires. In the examination surveys, the information is collected using questionnaires in combination with physical examinations, which often include the collection of blood samples and/or other specimens.

The cornerstone of any epidemiological inference, regardless of the data source, is the accuracy and the reliability of the data. The accuracy of the data refers to the degree to which the measurement represents the true value of the attribute being measured. Sometimes, also, the terms precision or validity are wrongly used as synonyms for accuracy. Precision of the

measurement only defines the resolution but does not necessary imply that the measurement is accurate, while the validity in epidemiology means that the test measures what it is supposed to measure. Reliability, also called repeatability or reproducibility, means that the measurement gives the same results when repeated. [1] The accuracy of the data does not imply that the data would also be reliable or precise and vice versa.

Using a survey for the data collection will create a number of potential sources of error due to the large number of steps needed for the completion of the survey. The *Figure 1* describes how the total survey error is divided into sampling and non-sampling errors. The sampling error is related to the coverage of the sample while the non-sampling error is related to the errors in specification, sampling frame, non-response, measurement, and the processing of data. [2, 5]

In relation to the survey error, the terms bias and variance are often used to describe the error. Bias refers to the systematic error, i.e. all the measurements systematically deviate from the true value. Variance is due to variable error, a random deviation in the measurements. [1, 5]

Standardization and quality control can improve the accuracy and reliability of the survey data. With standardization of the methods, the comparability of results between surveys can be increased. The purpose of standardization is to remove as many of the potential sources of error as possible [1]. This can be done with standardized measurement protocols and instruments, i.e. the same instruments are used in different surveys, and the measurements are taken using the same measurement protocol. Quality control will help to monitor the use of a given standards and to detect possible systematic or excessive random errors [1]. When systematic or excessive random errors are detected early enough, they can be corrected and the accuracy and reliability of the data will remain high.

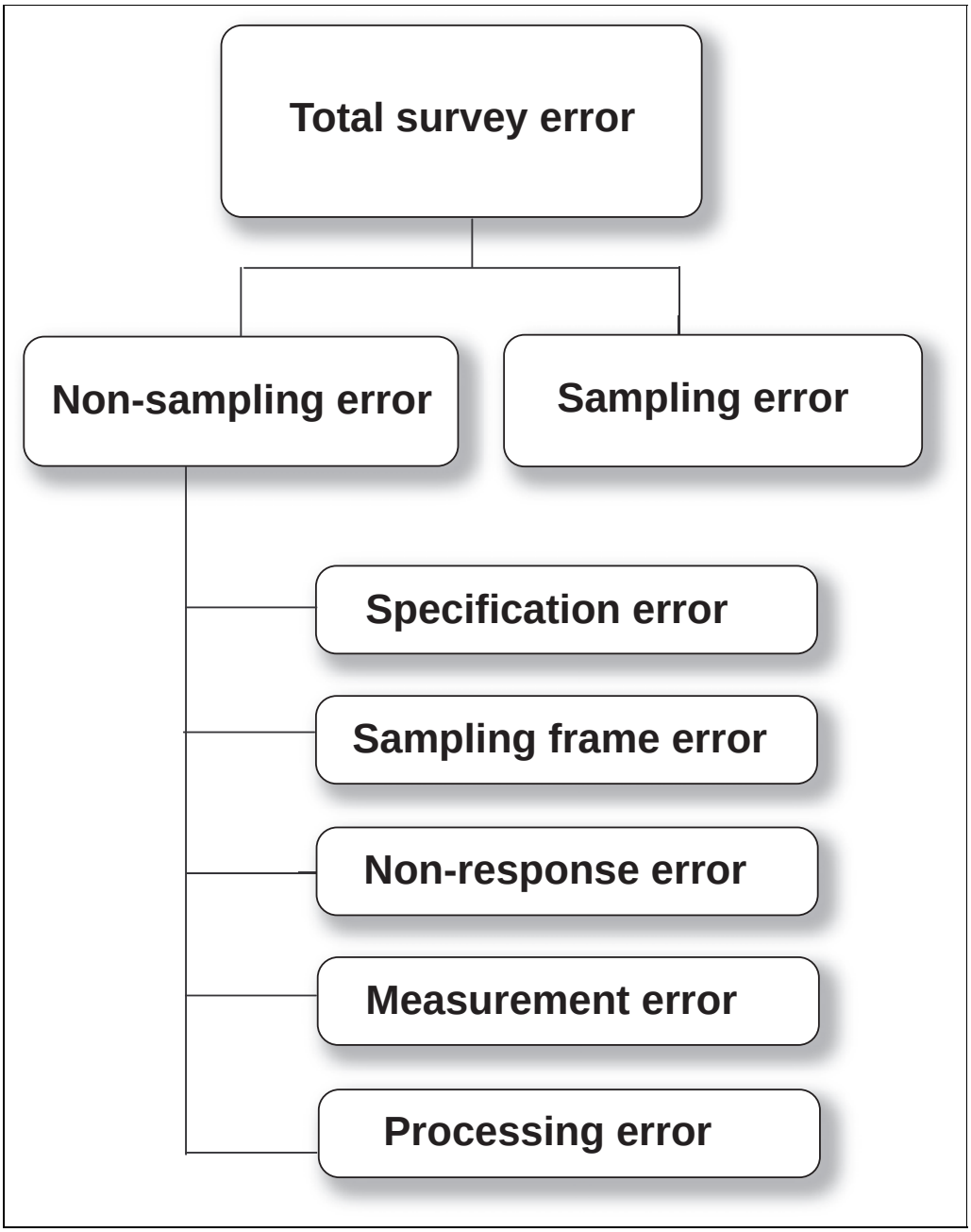

Figure 1. Components of the total survey error

This thesis got started with the work I did for the WHO MONICA Project [6, 7] preparing a number of retrospective quality assessment reports [8-16]. These reports assessed and summarized the achieved quality of data collected around the world using standardized protocols. Even though the methods of the MONICA risk factor surveys were standardized and there were very detailed instructions, all participating centres were not able to follow them in all parts. There were also biases due to problems related to the deviation from the standardized protocol.

Being closely involved with this process of assessing the data quality for a multinational study and later in drafting the proposal for the chronic disease risk factor survey methods for the European Union [17] got me interested in survey data quality and standardization.

The purpose of this work is to study the survey process from the perspective of data quality and the effects of data quality on the interpretation of the results. It covers the survey process from planning, to implementation and to reporting of the results. The papers written for this thesis deal with different components of the overall data quality and demonstrate how the changes in methods, low response rate, data quality, and the indicators used for reporting may affect the results.

2 LITERATURE REVIEW AND DISCUSSION

In the literature, the term survey is generally used to describe studies where information is collected using questionnaires. Therefore, the survey literature in most cases covers only the methods relating to the questionnaire based surveys [18-26]. There is very limited survey literature, which also discusses issues relating to the data collection by physical measurements, including blood samples and other specimens, in a survey settings [17, 27-32].

This imbalance of the survey literature between the questionnaire methods and the methods relating to physical measurements is understandable. Same questionnaire survey methods that are used in health surveys are also applicable for the sociological, economical and many other surveys. As the questionnaire is the common tool for different kinds of surveys to collect information, it is natural that most of the research and development of the survey methods has been focused on this area. Physical measurements are used only in a limited number of health-related surveys and therefore, research and development of these methods is not as extensive as of the questionnaire survey methods.

In this literature review, the survey process is covered with respect to data quality. The *Figure 2* described the order in which the survey process is reviewed here, starting from the definition of the survey objectives and then moving to the selection of survey mode and sampling, including the decisions about the sampling frames, sample size and the actual sampling methods. The issues relating to the planning of the survey logistics and the selection of the personnel are reviewed. Following, the issues relating to the questionnaire design and the selection of the questionnaire administration mode are examined. As the main focus is on health surveys, the selection of physical measurements is discussed and the preparation and contents of the survey manual is reviewed. Finally, the review includes issues relating to the training of the personnel, quality control during and after the survey (retrospective quality assessment), and to the reporting of the results. Documentation and data management issues are discussed at the end, even though they closely relate to each step of the survey process.

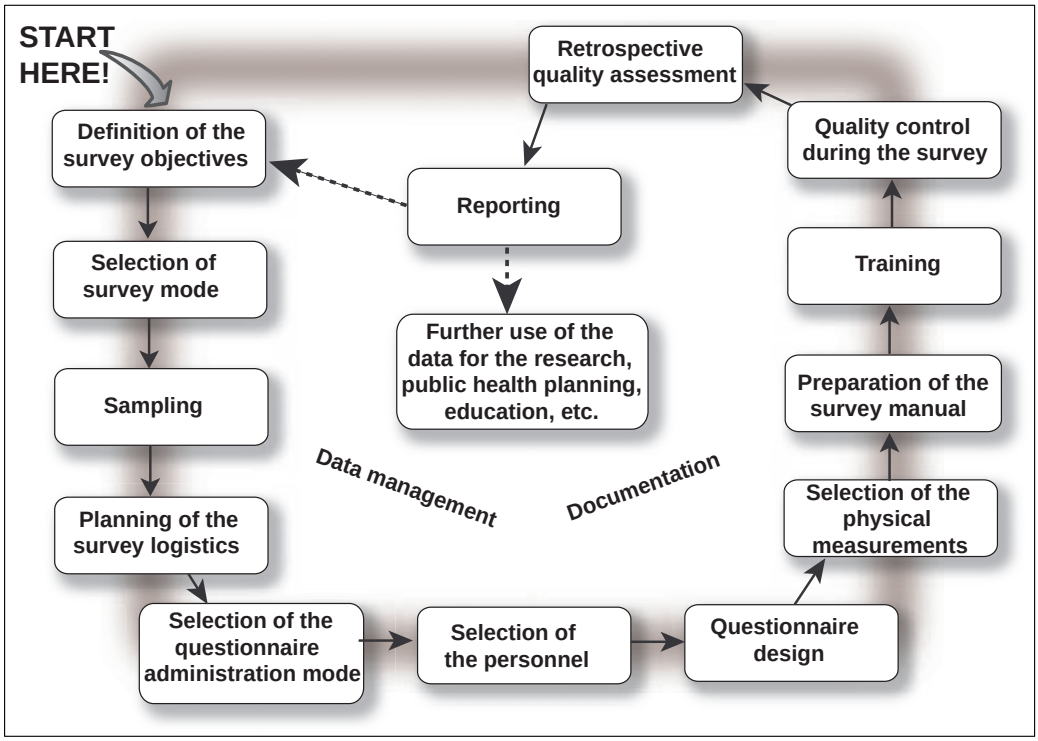

Figure 2. Order of the survey process

For all these steps, the literature on different available methods is reviewed and their possible implications to data quality are discussed.

2.1 Selection of the survey objectives

The health survey, just like any survey, is a huge undertaking on the part of the researcher to find the answers to the research questions. That is why the planning of the survey is the most essential part of the whole survey process. Good planning helps to obtain better outcome and saves both time and money during the survey.

Researchers usually have a number of research questions in mind when they start the planning of the survey. The objectives of the survey are defined through these research questions. It is important to be critical with the objectives already during the preliminary stage of planning.

Individual research questions should have some real evidence behind them. The argument *"It just would be nice to know this"* is not enough. Nowadays, the number of all kinds of surveys is increasing, and people are easily getting tired of participating in surveys after surveys. Therefore, it is important to think of the respondents' burden when selecting the questions to form the survey objectives.

After the rationale for the individual research questions is established, the feasibility of them for the survey setting needs to be assessed (*Figure 3*). First, already existing data sources are reviewed to see if the information would already be available and therefore the survey would not be needed [25]. There are situations, when the data for some parts of the research question would be available in existing data sources but to get the complete answer to the research question would require more information. In situation like this, the survey is justified.

Even when the data are not available from the existing data sources, a survey is not necessarily a feasible or best way of collecting the information (*Figure 3*). For example, the measurement procedures needed to obtain required data might be so sensitive or difficult that their use in a survey setting is not possible and therefore the survey as data collection method is not possible.

After each research question has be evaluated and found suitable for the survey setting, the actual construction of the survey objectives can begin. Individual research questions are grouped into larger units, which form a coherent set. For the final objectives of the survey, several of these sets can be combined if they have some linkage between them. It should be remembered that a survey should have a uniform objective and all loose pieces are left out. For example, for a health survey one does not include some unrelated questions about politics or market prices of consumer products, regardless of how interesting and suitable for the survey setting they would be.

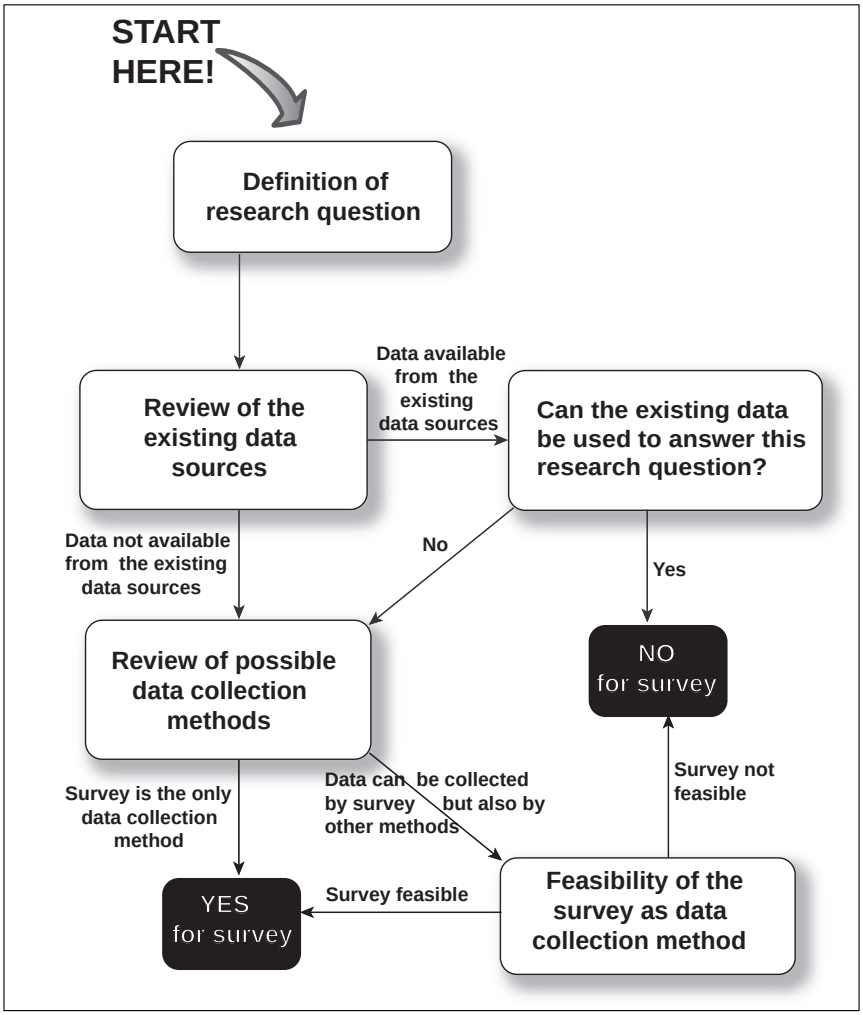

Figure 3. Evaluation process for the feasibility of the individual research questions for the survey setting

Setting up the final survey objectives requires balancing between available resources (funds, personnel, time, etc.) and research interests. Usually, the most limiting factor is money, which in the final stage will set limits to what can be done and how. This can also have an impact on the over all survey quality.

Example. Selecting one expensive measurement for the survey often means that several cheaper measurements have to be dropped. This decision will affect the amount of information obtained by the survey and therefore impact on the possibilities to answer survey objectives (*Figure 4.*).

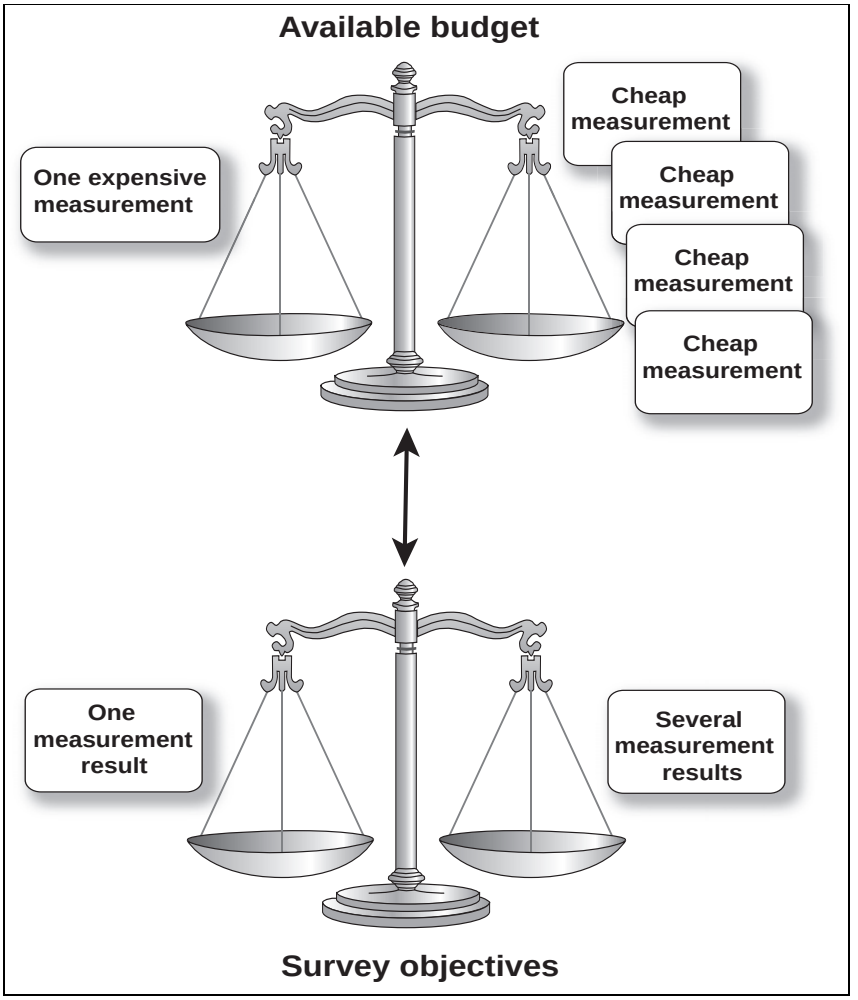

Figure 4. Balancing between survey objectives and survey measurements

Conclusions: To answer the survey objectives is compromising between available resources, research interest, and available methods.

2.2 Selection of survey mode

Survey mode means the tools used in the survey for data collection. There are two main survey modes; questionnaire surveys and surveys with physical examinations. Questionnaire surveys can be either interviews or self-administered, when respondents fill in the questionnaire themselves. Surveys with physical examinations generally include a questionnaire and a number of physical examinations, collection of blood samples, or other specimens. The field of physical examinations, which can be made in the surveys, is wide and growing all the time.

Terms like health interview survey (HIS), health examination survey (HES), risk factor survey, health behaviour survey, and health survey are commonly used to describe the different survey modes in relation to surveys measuring populations' health and the components of health. Health survey is a general term referring to all surveys, questionnaire, and examination surveys, which deal with the populations' health. Health interview and health behaviour surveys are questionnaire surveys. The health interview surveys usually refer to any questionnaire survey dealing with populations' health while the health behaviour surveys usually are more focused on the health behaviours of the population, like dietary and smoking habits. Health examination and risk factor surveys are surveys with physical examinations. The health examination surveys are general health surveys with physical examination, without any focus on specific disease or disease group. The risk factor surveys are health surveys, which usually are focusing on the risk factors of some disease or group of diseases, like cardiovascular disease risk factors.

The selection of survey mode has an effect on data collection tools, personnel, budget, turnaround time, requirements for the respondent, and the types of data that can be collected (*Table 1*).

Table 1. Characteristics of the questionnaire and examination surveys

Characteristics	Questionnaire surveys	Examination surveys
Data collection tools	Questionnaire	Questionnaire Measurement devises (example: scales, sphygmomanometers, etc.)
Personnel	Administrative staff Interviewers if interview survey Data management staff	Administrative staff Interviewers if interviewer survey Staff to conduct physical measurements and to collected blood samples/specimens Data management staff
Price	Low/medium	High
Turnaround time	Fast/medium	Slow
Requirements for respondent	Knowledge about the issues under study	Capacity to take part in physical examinations
Type of data	Behaviours Subjective health Diagnosed diseases	Biological status Diseases

When the questionnaire survey mode is selected, the only available data collection tool is the questionnaire. The need for staff is rather small, when the questionnaires are mailed to the respondents. If the questionnaires are filled in during the interview, the need for staff increases to include a relevant number of interviewers. The questionnaire surveys do not require any expensive data collection tools and especially when the questionnaires are self-administered, the requirements for personnel are small and also the price of these types of surveys is usually low. The use of interviewers will substantially increase the price of the survey. The turnaround time for the questionnaire surveys is considered to be fast when questionnaires are either mailed to the respondent or filled in during a telephone interview. If the personal interviews are used, the turnaround time will increase, as the interviewer has to

travel from place to place to conduct the interviews. Questionnaire surveys require from the respondent knowledge about the issues studied. Also, how the questionnaire is filled in; self-administration or interview will have its own effects on the requirements from the respondent (see Chapter 2.7.3. more details). Generally, the questionnaire surveys can be used to collect information about behaviours, subjective health, and diagnosed diseases. (*Table 1*)

The examination surveys offer a wider range of data collection tools than the questionnaire surveys, including a number of physical measurements in addition to the questionnaires. The expansion of data collection tools will also expand the need for personnel. Just like in questionnaire surveys, administrative and data management staff is needed and if the included questionnaires are filled in during an interview, a number of interviewers is required. In additional to that, each physical measurement will require trained personnel and a special place to conduct the measurements. This place may be a local health centre but still it needs to be prepared for the survey measurements, which will be different from general visits to the health centre. As the number of data collection tools and the required personnel increases, the price of the survey also increases. This makes the turnaround time slower, since the examination of each respondent will take more time than just filling in the questionnaire. From the respondent, the examination surveys usually require some physical and mental fitness to arrive to the survey site and to go through the examinations. The examination surveys can provide information about the biological status of the person and about the undiagnosed diseases. (*Table 1*)

Questionnaire surveys work well when the survey questions are simple enough to be answered without any additional physical examinations or tests and it can be assumed that the respondent knows the answers to the questions. Physical examinations are needed for example, if one is interested to know the population blood pressure profiles. This information cannot be obtained by questionnaire since one cannot expect everyone responding to the questionnaire to know their blood pressure. Also, querying administrative registers like hospital records will not inform about the population profile, only the profile of those hospitalised, and having had their blood pressure measured during their hospital stay.

There may be cultural differences in attitudes towards different survey modes. In developing countries, the questionnaire surveys are often found difficult due to linguistic problems. People are not familiar with the terminology used in the questions and therefore do not understand questions and cannot answer them. Also, the concept of time may differ from the traditional western concept of time, which causes recall bias. Many issues relating to health are culturally very sensitive in the developing countries. The biggest limitation for the questionnaire surveys in the developing countries is that people actually are not aware of their conditions or diseases. [33]

Earlier, the examination surveys caused some problems in the developing countries due to requirements of laboratory testing (cold chain), and other measurements. Lately, the development of measurement systems and the introduction of so-called dry-chemistry tests have made these tests easier also in countries without developed infrastructure. [33]

Conclusions: The selection of survey mode has an influence on what kind of data can be collected and, therefore, what kind of information will be available to answer the survey objectives.

2.3 Ethics and confidentiality

The World Medical Association Declaration of Helsinki defines the ethical principles for the conduct of medical research involving human subjects [34]. Population health surveys are medical research on people, and therefore should follow these principles.

Briefly, the Declaration of Helsinki defines the right of the subject and obligations of the researcher. The subject has the right to refuse to participate in the survey, i.e. the participation is voluntary. The subject also has the right to know the aim of the survey in which he/she is participating, the methods used in the survey; who is involved in the survey (sponsors, organizers, etc.) and, of course, the benefits or potential risks of the survey to the participant. [34]

The researcher is responsible for making sure that the survey does not harm the participant in any ways, and that all the medical measurements are conducted by qualified persons. The researcher also has to respect the privacy of the subject and make sure that the information obtained is stored and handled with confidentiality. Finally, the researcher is responsible for making sure that published results are correct and that negative findings are made also publicly available. [34]

Each research protocol has to be approved by a local ethics committee, which is independent of the research group and its sponsors and other connections and should also conform to the laws and regulations of the country in which the research takes place. From each subject participating in the research project the informed consent is needed. The informed consent has to be signed by a legally competent person. [34] Examples of informed consent can be found, from example, in the Cardiovascular Survey Methods, 3rd edition [35].

Since the participation in any research project is voluntary, the survey organizers have to remember that they cannot force anyone to participate and they have to respect anyone who does not want to participate. Everyone among the staff must know and understand the importance of the confidentiality of the responses. It is important to explain to the respondent that answers are treated with confidentiality but that they are not anonymous in a sense that the person can be identified from each questionnaire/measurement. [25, 35]

A survey organizer who does not respect the privacy of the people and the voluntary nature of participation and is too aggressive in recruitment will easily end up with a low response rate and also cause harm for other future survey organizers. If people feel that a survey had once violated their privacy, they will become sceptical and cautious about any future surveys even though those would be conducted discreetly and observe the right of privacy of people. [25]

Conclusions: Following ethical and moral rules will help to obtain higher response rate and more reliable results.

2.4 Sample selection

2.4.1 Definition of target population

Target population is the population to which the inference is drawn by the survey [2, 23]. The general characteristic of the target population are: i) the target population is finite in size, 2) it exists within a specified time frame, and iii) it can be accessed [23].

> *Example.* The target population could be the inhabitants of the city of Helsinki. Inhabitants of Helsinki can be counted, therefore being finite in size; the inhabitants change over time when people are moving in and out, die, and are born, therefore the target population exists within a specified time frame. The contact information for the inhabitants can be obtained through a population register; therefore, the target population is also accessible.

Sometimes availability of a good sampling frame (a list of sample units in the target population) may limit the selection of the target population. For example, if the only available sampling frame for the planned target population covers only a certain age group, the target population will need to be limited to that age group.

For practical reasons, some population groups may be excluded from the target population, even though they would fulfil the criteria set for the definition of target population. The criterion used to define the target population is called eligibility criterion [1].

> *Example.* The target population is the inhabitants of the city of Helsinki, but hospitalised or otherwise institutionalised people are classified as ineligible for the survey. This limitation is imposed because those people are often physically and/or mentally incapable to provide accurate and reliable information for the questionnaires and to take part in the physical examinations.

This eligibility criterion for the target population will set limitations for the collected data. Hospitalised and institutionalised people are still part of the inhabitants living in the city of Helsinki. Their health profile most likely is very different from the health profile of the general population. Leaving these people out from the target population will provide results, which are only representative for non-hospitalised and non-institutionalised inhabitants of the city of Helsinki. One does not have any information about this fraction of the inhabitants who usually consume the most of the health care services of the city of Helsinki (hospitalised and institutionalised) and therefore our estimates of required health care services based on obtained data will be under estimated.

Conclusions: The definition of target population will affect the representativeness of the results.

2.4.2 Age group to be studied

A part of the definition of the target population is the selection of the age group to be studied. The objectives of the survey will define the age group to be studied but the available budget and existing sampling frames may set some limits.

Example. The interest is to study the health profile of the entire population but due to the limited budget, one has to restrict the survey to the age group of 18-74 years.

Example. The interest is to study the health profile of the entire population but the available sampling frame lists only the adult population aged 18 years or more. Therefore, one has to limit the survey to the adult population of age 18 years or older.

The selection of the age group covered by the survey will affect the conclusions that can be drawn from the results, the comparability of results with other similar surveys, both nationally and internationally, and the survey methods to be used. If the age group surveyed is 18-74

years, the conclusions can be drawn only for that age group or some smaller age group within the range of 18-74 years. Also, the comparisons with other surveys are limited to the age range of 18-74 years. For small children often proxy respondents are needed to provide information about their health behaviours and health status. Children also often require questionnaires and measurement procedures different from adults.

When ever possible, the target age group should be selected to conform to the common commonly used age group categories. The Medline [36] uses the age groups:

- *All infants:* birth – 23 months
    - *Newborn:* birth – 1 month
    - *Infant:* 1 – 23 months
- *All children:* 0 – 18 years
    - *Preschool children:* 2-5 years
    - *Children:* 6 –12 years
    - *Adolescent:* 13 – 18 years
- *All adults:* 19+ years
    - *Adults:* 19 – 44 years
    - *Middle aged:* 45-64 years
    - *Middle aged + aged:* 45+ years
    - *Aged:* 65+ years
    - *80 and over:* 80+ years

The tenth revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) [37] specifies three different groupings, which are not compatible with those in Medline:

- Under 1 year, 1-4 years, 5-year groups from 5 to 84 years, 85 years and over;
- Under 1 year, 1-4 years, 10-year groups from 5 to 74 years, 75 years and over;
- Under 1 year, 1-14 years, 15-44 years, 45-64 years, 65 years and over.

Change in age structure of the population in many countries will set new challenges for the selection of the age group to be studied. Countries may have long traditions to survey adult (working age) populations but now when people are living longer there is pressure to include the elderly in the surveys.

Conclusions: The selected age group will affect the selection of survey instruments and methods. The selected age group also affects the representativeness and comparability of the results.

2.4.3 Sampling frame

Sampling frame is a list of sample units in the target population. A good sampling frame includes all population units while excluding any non-population units, does not have any duplicates, assigns units to the correct cluster (age, educational level, etc.), has correct auxiliary information, and has up-to-date contact information. [2]

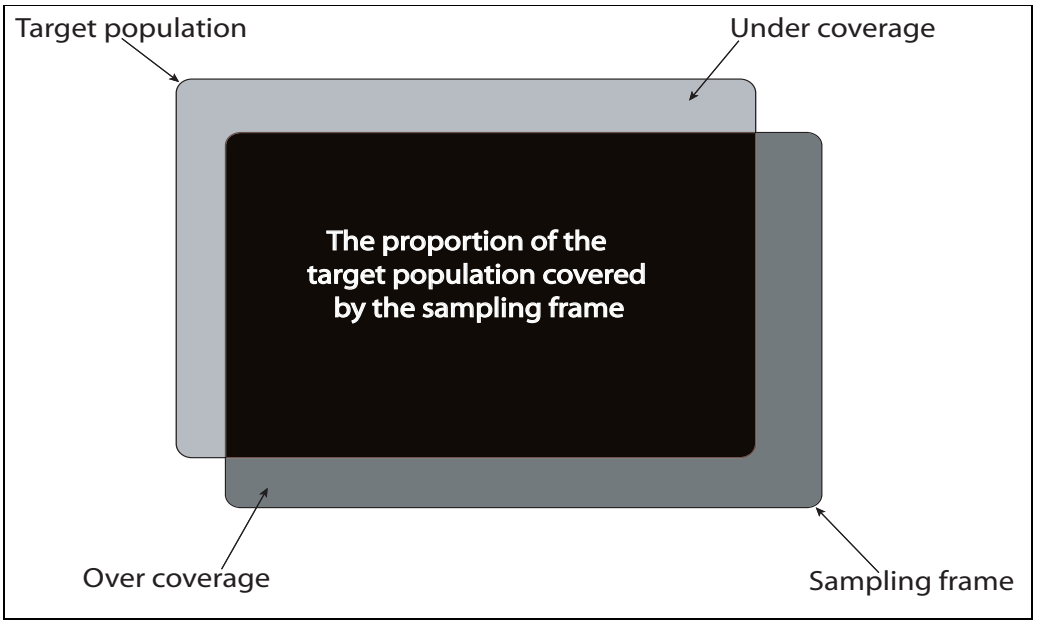

Figure 5. Coverage of the target population by the sampling frame

The difference between the target population and the actual population covered by the sampling frame may create a coverage error of the survey [25] as shown in the *Figure 5*. Missing population elements in the sampling frame cause under-coverage, also called non-coverage or incomplete coverage. These missing population elements are usually hard to detect and they will cause under-estimation of totals and possibly will also bias other survey statistics. Over-coverage is caused by inclusion of non-population elements in the sampling frame. These non-population elements can usually be detected during the survey and be removed. If they are not removed from the sampling frame, they will cause over-estimation of totals and possibly will also bias other survey statistics. A multiplicity problem is caused by duplicate listings of sampling units in the sampling frame. Duplicate listings should be removed or proper sampling probabilities based on the number of times persons appear on the list should be used. Incorrect auxiliary information is caused by deviation in contents, i.e. information may be in error or missing for some frame units. Incorrect auxiliary information will decrease the precision of the survey information. Incorrect access information, i.e. out-of-date or inaccurate address can sometimes be recognized in the frame. This is a case when one can identify each member of a sampling frame, for example former students of a school, but

do not have up-to-date contact information for some of them. [2] Often, the recognition of incorrect access information may be difficult. When the survey questionnaires or invitation letters to the examinations are mailed to the survey participants and some do not reply, one cannot always confirm that the reason for no reply is the refusal to participate or not receiving the questionnaire/invitation letter.

The selection of the sampling frame has to be based on objectives of the survey. For example, if the target population for the survey is the population of diabetics, it is not feasible to draw the sample from the population register since the information about the diabetes status is not available there. On the other hand, drawing the sample from the member list of local diabetic association will not result in a representative sample for the entire population.

The number of different types of sampling frames is large and their availability varies from country to country and sometimes event within a country. A few commonly used sampling frames are described below. This list of sampling frames is not complete and will change over time when new lists are created and old ones are terminated.

Conclusions: Coverage of the target population affects the representativeness of the results for the target population.

2.4.3.1 Population registers and census

National population registers list each individual living in the country. Each individual has a unique identification code that incorporates a check digit. In these registers, the correspondence between the resident population and identification codes is 1:1. It is obligatory to inform registration authorities about the change of address and the registration authority receives information about births, deaths and marriages from corresponding authorities directly and promptly. Therefore, population registers are well up-to-date. [38]

National population registers exists at least in Scandinavia [9, 38-40], the Baltic countries [41-43], Belgium [9, 38], Luxembourg [38], the Netherlands [38, 44], and Portugal [38] .

Local or regional population registers exist at least in Germany [9, 38], Greece [38], Switzerland [9], Spain [9] and Italy [9, 38].

Population census, i.e. enumeration of the entire population at a fixed time is organized in the most countries on a more or less regular basis. Since the organization of a census is a huge undertaking, therefore, a census is usually organized only every ten years or so. [45-47] In many countries, the census means approaching each household either by mail or by personal visit [45, 46] to obtain required information. In some countries, the census is conducted as register-based census. This is the case for example in Finland, were a national population register exists with each person having a unique identification code. Using these identification codes, several administrative registers can be linked and the population census conducted without actual enumeration of the population. [47]

Since the census is conducted infrequently, its coverage is good only at the time of the census and shortly thereafter but soon it gets out-of-date [48].

National and regional population registers are ideal sampling frames for population samples when they exist and are available for such use. The usability of the census as sampling frame is limited to the time when census occurs. Both, the population register and the census can provide a fairly complete list of the population in all age groups.

2.4.3.2 Telephone directories

Telephone directories are often used as sampling frames for surveys that use telephone interviews. The availability of information about the telephone coverage in households is very limited, since this is not available from any administrative registers. In Finland, 99% of households have either fixed-line or mobile phone [49]. This is also true for Great Britain [50] and Germany [51]. In the USA, the telephone coverage in households is a little lower, 94% [52]. For many countries, this information is either not available or not published.

Table 2. Telephone coverage in 2003. Number of telephones per 100 inhabitants (variation in brackets)

Continent	Telephones (fixed and mobile) per 100 inhabitants	Fixed-lines per 100 inhabitants	Mobile telephones	
			per 100 inhabitants	% from all telephones
Europe	96 (24-186)	41 (8-80)	55 (8-106)	58% (27%-81%)
Africa	9 (0-82)	3 (0-29)	6 (0-68)	67% (11%-99%)
Americas	67 (3-116)	34 (2-63)	34 (0-54)	50% (3%-87%)
Asia	29 (1-170)	14 (0-59)	15 (0-111)	52% (2%-92%)
Oceania	95 (2-126)	41 (1-54)	54 (0-71)	57% (12%-63%)
World	40	19	22	54%

The statistics by the International Telecommunication Union can help to get an idea of the variation in telephone coverage in general (household and business telephones) between countries. These statistics report on an annual basis the number of telephones per 100 inhabitants in each country and they also provide summary statistics by continent [53-55]. For example, in 2003, there were in Europe on average 96 telephones per 100 inhabitants while at the same time in Africa the number was 9. The variation within continents was even more dramatic, especially in Asia. (*Table 2*)

The International Telecommunication Union statistics report the number of telephone lines per 100 inhabitants. This figure will overestimate the telephone coverage in households, since these figures include also office telephones. From these statistics, one cannot get information about the telephone coverage in different population groups within a country.

From available household survey results, the differences on the socio-economic status of telephone owners can be found. In the USA, people without telephone tend to belong to the lower socio-economic class [56] and are more often African Americans [57]. In Great Britain, people

without telephones tend to be older, single, belong to lower-income groups and are unemployed [58]. Also in Finland, people without telephone are more often single and older men [59].

One problem with telephone directories as sampling frames is the group of unlisted numbers. For example, in the USA already in mid-1980's 20% of telephone numbers were unlisted and the proportion was higher in big cities like in New York (25%) and in Los Angeles (34%). At the same time, in the Netherlands only 4% and in the Great Britain 12% of telephone numbers were unlisted. [58] Ten years later, at the end of the 1990's, in the Great Britain, 34% on telephone numbers were unlisted and in the London area, the proportion was as high as 44% [60]. Finland still has a relatively small proportion of unlisted telephone numbers (9%) but there is a geographical variation in that, the capital area has relatively more unlisted numbers than other parts of the country [61].

This increase in unlisted telephone numbers has an effect on the usability of telephone directories as sampling frames. When more and more people have unlisted telephone number and telephone directories are used as sampling frames, the representativeness of the sample for the target population decreases. For example, in the London area 44% of telephone numbers are unlisted. This also means that 44% of target population would be excluded from the sample if the telephone directory would be used as sampling frame. Exclusion of this magnitude will cause significant bias to the results.

In many countries, the structure of telephone coverage has been changing due to mobile phones. While the number of mobile phones has increased, the number of fixed-line telephones has decreased. For example, in Finland, the proportion of households with mobile phones in 1996 was only 7% and by 1999 rose to 20% reaching 87% by year 2001 [49, 59, 61]. In Finland, the proportion of households with mobile phone exceeded the proportion of households with fixed-line telephones in 1999. Presently, in almost half of the households there is more than one mobile phone. [59] The proportion of households with mobile phones is also high in Great Britain and Germany. (*Table 3*)

Table 3. Ownership of different kinds of telephone lines

Country	Year	Household telephone coverage	Fixed-line telephones	Mobile telephones	Reference
Finland	2001	99%	69%	87%	[49, 59]
UK	2002	99%	92%	75%	[50]
Germany	2003	99%	95%	73%	[51]

The increase of mobile phones has not been uniform throughout the population. At least in Finland, mobile phones are more common among young (19-29 years of age) and single people who live in cities [49].

Telephone directories are good sampling frames, if the country has high coverage of telephone lines, low proportion of unlisted numbers and if telephone directories are updated on a regular basis. In a growing number of countries, mobile phones are replacing the traditional fixed-line telephones. This will affect the use of telephone directories as sampling frames. In case of fixed-line telephones, lines are associated with a household and can be used to select household sampling units, whereas mobile phones are personal and are thus suitable to select individuals.

2.4.3.3 General practitioners listings

Frequently used sampling frames are general practitioners listings, especially in countries where population registers do not exist. For example, in Great Britain, every person is expected to register with a local general practitioner (GP) but the registration is not compulsory.

Still in the mid 1990's, the GP lists in Great Britain were found to have coverage problems. GP lists tended to include dead people, outdated addresses and wrong birth dates [62].

Recently, the National Health Service has developed the system and due to changed payment policies, the accuracy of GP lists has improved.

GP lists are usable sampling frames if they are frequently updated as they cover the entire population from infants to elderly. There are always some people who do not register with a GP and therefore will be missed. If it is known, that certain population groups like the homeless are more often not listed than others; they should be given special consideration while planning the use of a GP list as sampling frame.

Even though the example here is about the GP lists in Great Britain, other countries also have GP lists and other health service registers [9], which can be used as sampling frames.

2.4.3.4 Voter and electoral registers

In some countries, there are special registers listing people eligible for voting in elections. Such registers exist for example in Australia [9], France [9], Italy [9], New Zealand [9], Poland [9], Great Britain [63] and the USA [64]. In Great Britain, a person has to register with an electoral register if he/she wants to take part in elections. Persons are eligible for voting if they are 18 years or older, and are British, Commonwealth or European Union citizens. One can register to electoral register already at age of 16. [63]

In the USA, a person registering with a voter register has to be a citizen of the USA. Age of registration in general is 18 years but there are differences between States when person should turn 18 in relation to the date of the election and registering with the voter register. Also, there are differences between States with respect to eligibility criteria. In some States, convicted felons are not eligible and in other States persons declared mentally incompetent can not vote. [64]

As can be seen from the above two examples, voter and electoral registers are limited to the voting age population and in many countries only to the citizens of that country. These registers are not updated automatically through administrative procedures but require that people themselves actively register.

Because voting and electoral registers are not updated routinely when people move, die, or reach the age when they can vote, these register tend to have a lot of invalid addresses and missing information. It has been shown that for example in the Great Britain, 4-7% of the eligible population are not listed in the electoral register and 38% of total population is missing from the lists because they are non-British or non-Commonwealth citizen [38, 48]. Especially, for the elderly, the electoral registers tend to be badly out-of-date, listing only 39% of elderly citizens [65].

In the USA, voting registers cover 63%-70% of the entire population but the coverage varies by age and ethnic groups [57, 66]. For example, during the presidential election in 2000, 8% of population could not register for voting since they were not citizens. From white non-Hispanics, only 2% are non-citizens while among blacks, the proportion is 6%, Asian and Pacific Islanders 25%, and Hispanics 39%. [66]

The available information in the electoral and voter registers may vary from country to country. For example, in the USA, the voter register has information about the name, address, date of birth and the race or ethnic group to which person belongs [64] while in Great Britain the electoral register only has information about the name and address of the person [63].

In some countries, the electoral and voter registers may be the only available population lists and therefore the best available sampling frame. The limitation of these registers is that they cover only the voting age population and in many cases only citizens of the country, which will create coverage bias if the target population is the entire population of the country.

2.4.3.5 List of properties and other population lists

Population registers, telephone directories, GP lists, and voting registers are often used sampling frames. For example, in the WHO MONICA Project, the national or regional population registers were the most commonly used sampling frames, together with voting

registers and different GP and health service listings. In some populations also census lists, and different kinds of household lists served as sampling frames. [9]

There are an unlimited number of different kinds of list and registers, which have population information. Many times these lists and registers are limited to a specific population group but they still can be used as sampling frames if nothing better is available.

For example, in the USA driver's licence lists have been found to have good coverage of the population aged 15-74 years of age (90%) [67] but the address information tends to be out-of-date [57]. In comparison to the voter registers, the driver's licence registers are better up-to-date and they list more African Americans than the voter registers. However, in 1994 the USA outlawed the use of driver's licence lists as sampling frames. [57]

Also, different kinds of utility and property owner lists are available in many countries. In Great Britain, all property owners are listed in the rating valuation lists. These lists are well up-to-date and have good coverage at the local level but central lists tend to suffer from delays in updating. [48] Corresponding lists are available also in the USA [57]. The limitation of these lists is that they contain only properly owners' not actual people living in the houses.

One commonly used type of the sampling frames is the different kinds of address lists. These address lists usually list each house on the street/area. For example in Great Britain, the postcode address file (PAF) lists each private household address. This list is well up-to-date, listing 98% of the addresses but it does not have information about the people living at the address. [48]

2.4.4 Sample size

When deciding about the sample size for the survey, researchers should keep in mind that sample size has to be big enough to give reliable and precise results for the target population. Increasing the sample size will reduce the sampling error but will not reduce the possible measurement bias [68].

Balancing between optimal and feasible sample size is one of the tasks of the researcher [4, 69]. The amount of work during interviews and examinations and available funds will restrict the sample size. (*Figure 6*)

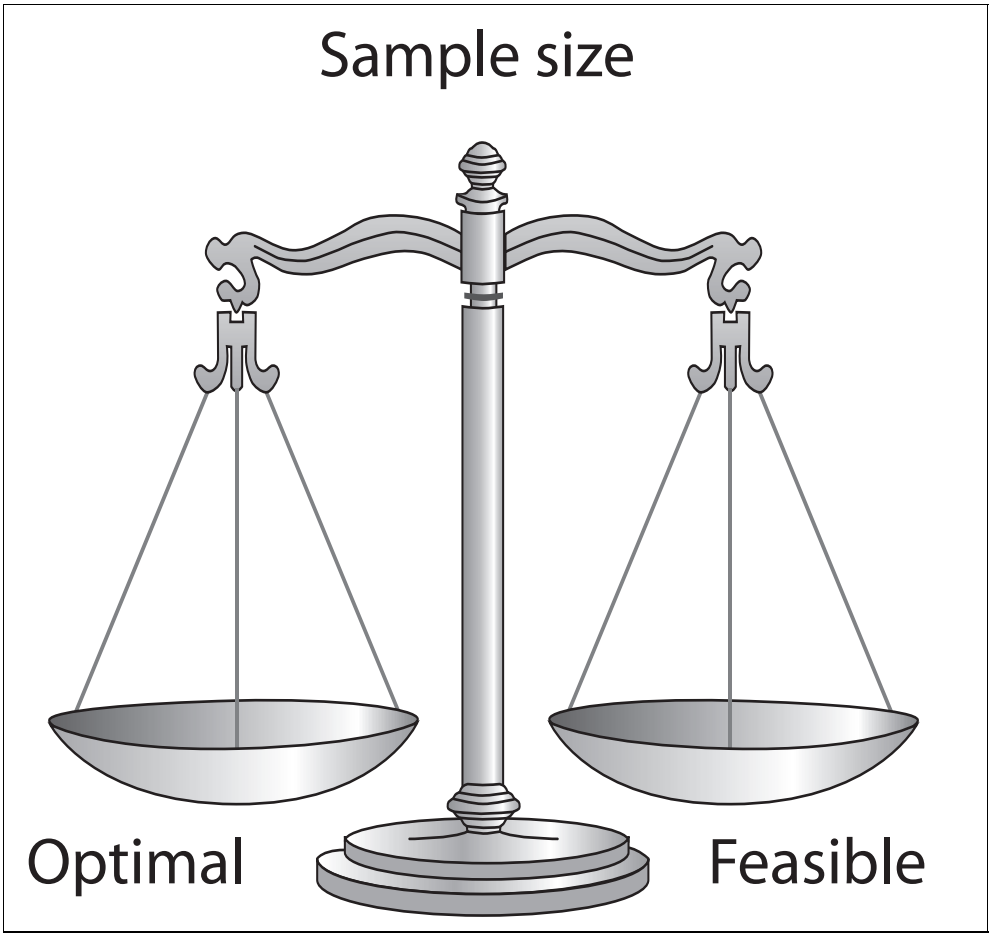

Figure 6. Balancing between optimal and feasible sample size

The required sample size can be determined using sampling theory. Depending on the quantity of interest (population mean, prevalence in population or differences between populations) there are statistical algorithms for sample size determination.

The basic idea of sample size determination for population mean (continuous variables) and prevalence (proportions) is the same. Information about the expected variance or prevalence is needed, either from previous surveys or from a pilot study. The level of precision (α), i.e. the amount of error that can be tolerated for the estimates and the allowed variation (*d*), i.e. the margin of error from the expected mean/prevalence has to be defined. (*Figure 7*) Based on this information, the required sample size can be calculated. [69]

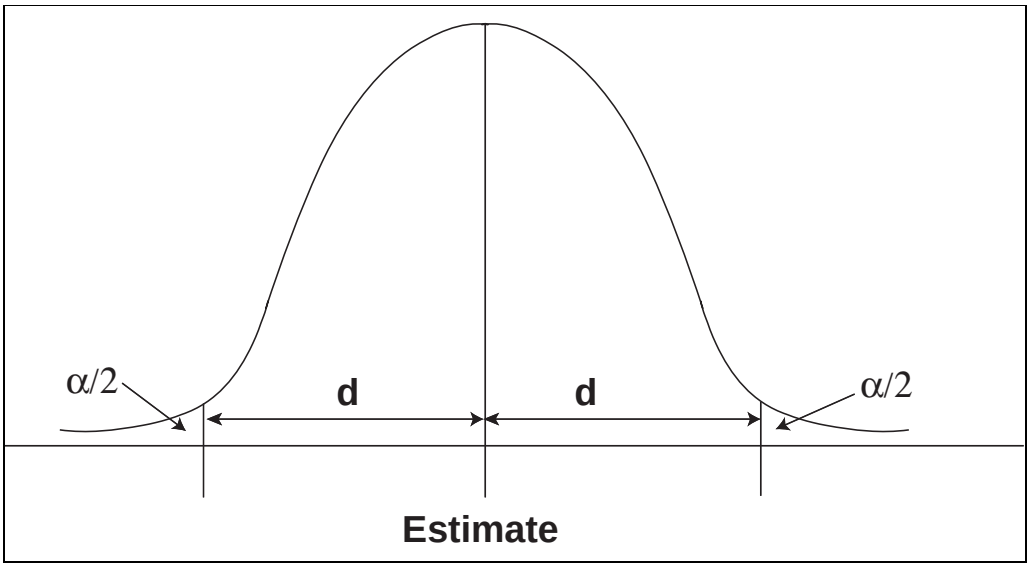

Figure 7. The level of precision and the allowed variance of estimate

When the main interest is to measure, for example, the prevalence of daily smoking in the population, the sample size can be estimated by equation:

$$n = \frac{z_{1-\alpha/2}^2 \times P \times Q \times N}{d^2 \times (N-1) + z_{1-\alpha/2}^2 \times P \times Q}, \quad (2.1)$$

where z_α is the value from the normal distribution at the α-cut-point, P is the assumed probability of the phenomena in the population, Q = (1-P), N is the population size and *d* is

the allowed margin of the error of the estimate. This model assumes that P is normally distributed. [69, 70]

When the population size is large, the equation (2.1) can be simplified to [69, 70]

$$n = \frac{z^2_{1-\alpha/2} \times P \times Q}{d^2}. \qquad (2.2)$$

Example. One has a population of size 5.000.000 inhabitants and one would be interested to know the prevalence of daily smoking in this population. One would like the population estimate to be precise at the 5% level (α) and have maximally a 5% error (d). From a previous study conducted in the same population 10 years ago one knows that smoking prevalence was 38% but one also knows that smoking prevalence has been declining in this population on average 0.7% per year. Therefore, one guesses that smoking prevalence would now be around 31%.

The required sample size in this case would be

$$\frac{1.96^2 \times 31 \times 69}{5^2} = 329.$$

If one is interested to know smoking prevalence by sex and age groups in the populations, the sample size has to be determined separately for each subgroup of interest. For example, in case were one is studying men and women separately for age group 25-64 by 10-year age groups, one has eight subgroups, and for each subgroup one would need a sample of 329 persons for a total of 2632.

Correspondingly, the algorithm for population mean is [71]:

$$n = \frac{1}{\frac{d^2}{z^2_{1-\alpha/2} \times \sigma^2} + \frac{1}{N}}, \tag{2.3}$$

where σ^2 is the variance of the estimate of interest.

If the population size N is large, the equation (2.3) can be written in the form

$$n = \frac{z^2_{1-\alpha/2} \times \sigma^2}{d^2}. \tag{2.4}$$

Example. One has a population of size 5.000.000 inhabitants and one is interested to know the population mean total cholesterol. One would like the population estimate to be accurate at the 5% level (α) and have a maximal error (d) of 0.1 mmol/l. In a previous study conducted in same population, the standard deviation of the total cholesterol was 1.1 mmol/l.

The required sample size in this case would be

$$\frac{1.96^2 \times 1.1^2}{0.1^2} = 465.$$

Again, if one is interested to know the population mean total cholesterol by subgroups, the sample size has to be defined separately for each subgroup (stratum).

The information about the needed population estimates for prevalence and standard deviation of measure of interest can be obtained from other studies conducted in the same population or alternatively from a pilot study.

The interest of the study may also be the differences or relationships between populations or population groups or changes in time within population. In two sample setting, the determination of sample size is very similar to the setting of one sample presented above. First the precision level (α) and the allowed variation (d) is defined. For the sample size based on prevalence, the expected prevalence in two samples is defined. For the sample size based on differences in means, the expected variance (equal for two samples) is defined. This information may be obtained from pilot studies or as prior information from the populations. [70]

For the sample size based on the prevalence, the algorithm is

$$n = \frac{z_{1-\alpha/2}^2 \times \left(P_1 Q_1 + P_2 Q_2 \right)}{d^2}, \tag{2.5}$$

where P_1 is the prevalence in the first sample, $Q_1=(1-P_1)$ and P_2 is the prevalence in the second sample, and $Q_2=(1-P_2)$ [70].

> *Example.* One wants to study the change in the prevalence of daily smoking in the population. It is assumed that the smoking prevalence during the first survey is 55% and during the second survey 40%. The level of precision is set to be 5% and the allowed variation to be 10%.
>
> The required sample size in this case would be
>
> $$n = \frac{1.96^2 \times (0.55 \times 0.45 + 0.40 \times 0.60)}{0.10^2} = 187.$$

The sample size based on the difference in means is estimated by algorithm

$$n = \frac{z_{1-\alpha/2}^2 \times \left(2 \times \sigma^2 \right)}{d^2}, \tag{2.6}$$

where σ^2 is the estimated variance of the means [70].

Example. One wants to study the change in total cholesterol in the population. It is known from previous studies that the standard deviation of the total cholesterol is 1.1 mmol/l. The level of precision is set to be 5% and the allowed variation to be 0.3 mmol/l.

Now the required sample size would be

$$n = \frac{1.96^2 \times (2 \times 1.1^2)}{0.3^2} = 103.$$

The WHO MONICA Project recommended that the sample size should be at least 200 for each stratum of interest. [28] In case of the WHO MONICA Project, where the interest was in 10-year age/sex groups this meant that sample size should have been at least 1200 (for age group 35-64) or 1600 (for age group 25-64). This recommendation was based on the calculation where the precision level was 5% and the change in daily smoking was from 40% to 60%, the change in diastolic blood pressure by 3 mmHg, and in total cholesterol by 0.3 mmol/l [68].

There are a number of sample size tables available [70], which give the required sample size for different situations and also many web sites which offer sample size calculators. These can be used during the planning stage of the survey, but local statistician familiar with the planned survey setting and the target population should always check the final sample size calculations.

Conclusions: Balancing between optimal and feasible sample size. The sample size will affect the reliability and precision of the results.

2.4.5 Sampling methods

Sampling methods can be divided into probability and non-probability sampling. In epidemiological risk factor surveys, probability samples should be used to ensure that the sample is representative for the target population. Only in circumstances where it is impossible to conduct probability sampling can non-probability sampling be considered.

The literature of sampling techniques is extensive [69, 71-73]; therefore only a brief description of different sampling methods is given here.

Conclusions: The sampling method will affect the representativeness of the sample.

2.4.5.1 Non-probability sampling

Non-probability sampling may be used if no sampling frames are available for probability sampling.

One example for non-probability sampling is *quota sampling* where a target population is divided into strata and within each stratum a number of people are selected to the sample based on a defined sampling scheme [24]. For example, each village may be a stratum and from each village, the survey personnel select a certain number of men and women in the defined age range.

When non-probability sampling methods are used, the precision (standard deviation) of the results cannot be estimates [69]. This means that results obtained from the non-probability samples are just point estimates without any knowledge about the variation.

2.4.5.2 Probability sampling

In probability sampling, everyone has a known, non-zero probability to be selected to the sample. It has been reported that when every one has equal chance to be selected to the sample the response bias will decrease [25].

In *simple random sampling,* every one in the sampling frame has equal probability to be selected to the sample. Each unit in the sampling frame is numbered from 1 to N. Then, as many random numbers between 1 and N are selected as is required to fill the sample size. [69] Random numbers can be taken from random number tables or generated by computer programs.

Random sampling can also be conducted with replacement when each selected unit is returned to the sampling frame before the selection of the next unit. In this case, the same unit may come to the sample more than once. [69] In population health surveys, sampling is usually done without replacement.

In *stratified random sampling*, the population listed in the sampling frame is divided into subgroups, strata, by some know factor, for example sex, age or geographical region. These subgroups are unique, i.e. they are not overlapping, and together they cover the whole population listed in the sampling frame. The sample is drawn independently from each stratum using any of the sampling techniques. [69]

The benefits of stratified random sampling are that one can guarantee to have big enough samples in each subgroup for precise population estimates. Stratified random sampling also allows to use different sampling methods for different subgroups if needed and also to obtain different sample sizes for different subgroups. [69]

> *Example.* One is conducting the population health survey in Finland. The main interest is the health of Finns but one also wants to study the health of Lapis in more detail. The sample is selected using stratified random sampling, where the primary sampling units are the 12 provinces of Finland. For the province of

Lappi one draws a relatively larger sample (over-sampling) than for the other provinces to allow the study of the health of Lapis from a big enough sample. When one recalculates the estimates for the whole of Finland, the results are adjusted to take into account the bigger sample from the province of Lappi.

In *systematic sampling*, persons listed in the sampling frame are numbered from 1 to N. Then the list is divided into *k* units, where *k* is the size of the total population listed in the sampling frame divided by the sample size (*Figure 8*). The first person selected to the sample is drawn from the first *k* units and after that one person from each k^{th} unit is selected. [69, 73]

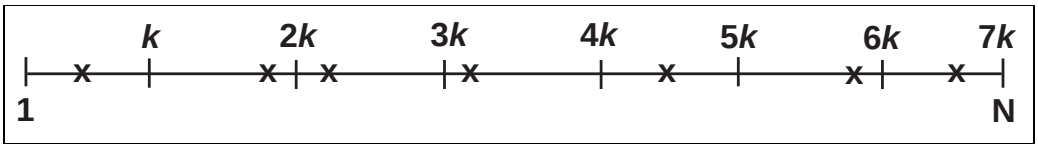

Figure 8. Systematic sampling

Systematic sampling is a good alternative when selecting a sample from telephone or other registers as well as from list of households. [73]

Sampling can also be conducted in more than one stage [71]. Multistage sampling is generally used due to practical reasons. Complete and up-to-date sampling frames covering entire population may be missing or expensive to obtain. In these cases, multistage sampling will provide the best possible sample.

Example. The population health survey is planned for Spain that does not have a national population register but has municipal population registers. Because of the lack of the national population register, the multistage sampling is used for the sample selection. In the first stage, the municipalities are listed. From this list, *x* number of municipalities are selected to the sample. In the second stage, in each of the selected municipalities, the sample of x_i is drawn from the municipality's population register.

In multistage sampling, sampling at each stage can be conducted using any of the existing sampling methods.

2.5 Survey logistics and selection of personnel

Survey logistics includes the planning of survey time schedule, selection, and organization of survey sites and all the issues related to the transportation of people and/or material from place to place during the survey.

2.5.1 Timing of survey

The proper timing of the survey is essential to reach as many people of the sample as possible and also to obtain accurate measurements for the indicators, which may be time dependent. [17]

When the survey concerns the working-age population, the timing of the survey should be planned so that those with regular day-time jobs as well as those on shift work can take part. For many working people it may be difficult to take time off work to participate in the survey, thus the schedule should allow for day time, evenings or weekends participation. Organizing additional interviews or physical examinations also during evenings and weekends may increase expenses and require special considerations for other survey logistics like handling of blood samples after office hours. [17]

It has been recommended that telephone surveys should be conducted between 18:00 and 21:00 on weekdays and between 12:00 and 21:00 on weekends. If calls are placed during the day on weekdays, they will miss those at work and if placed after 21:00, people might resent it as undue invasion of their privacy and their privilege to relax and go to bed. Also weekend calls should not be made before 12:00 to allow people to sleep in late and have peaceful weekend mornings. [74]

An additional timing issue for surveys is that on public holidays and during the holiday season people may be more difficult to reach. Some public holidays may have an effect on the outcome of the survey results, as might be the case if the holiday is related to fasting or other activities, which require alteration of the normal pattern of health behaviours. For example, in Finland where the use of alcohol is very heavy during Midsummer, the alcohol use indicators might be overestimated if the survey is conducted right after Midsummer and the alcohol use is assessed on the basis of consumption during the previous week. A similar kind of under- or overestimation could be observed for many other public or religious holidays/festivals where fasting or use of certain food groups plays an important role.

When the survey includes physical measurements, which are time dependent or require long fasting periods prior to sample collection, the timing issues get even more complicated [17]. If, for example, 12-hour fasting is required before blood drawing for glucose measurement, it would be easiest to ask participants to fast over night and come to the survey site in the morning. In practice, this would mean that physical measurements could be done only during the morning hours. When the measurements are restricted to a short time interval during the day, the duration of the entire survey will be prolonged, since only a limited number of people can be examined during the restricted time period.

Several physical measurements are known to have seasonal variations. To minimize the potential effects of seasonal variation on the results, the surveys should be planned so that the examinations of all population subgroups are distributed evenly throughout the survey period. Also, any repeated surveys should be conducted during the same months as previous surveys to minimize the seasonal effects. [17]

Conclusions: Timing of the survey will affect the response rate and the actual measurement results, if they are time depended (for example has seasonal or diurnal variation).

2.5.2 Place of survey

Selection of the survey locale concerns mainly surveys where participants are asked to come to a specific survey site for interview and possible physical examinations [17]. But even when interviews are conducted at home or work sites, it is important to make sure that the place is suitable for that purpose by providing needed privacy and allowing the interview to be completed without interruptions or distractions.

Selection of the survey location depends on the measurements, their order, and timing. For measurements and interviews there has to be enough privacy. Also, the temperature of the rooms has to be comfortable enough, especially if measurements require persons to take off their clothes. Some measurements may have special requirements for the survey site, like when blood pressure is measured, the room should be quiet, and for weight measurement the floor should not have a carpet. [17]

The survey site should also be located so that participants can get there without difficulty, i.e. the survey site can be accessed using public transportation or other commonly available means of transportation. [17]

If the survey location is difficult to access, it will reduce the response rate, and if the place is noisy, blood pressure measurements will be difficult to be conducted accurately.

Conclusions: Place of the survey will affect the response rate and possibly also the accuracy of some measurements.

2.5.3 Selecting the personnel

2.5.3.1 Required personnel

The number and type of personnel needed depends on the type and size of the survey [35]. When deciding on the amount of personnel, allowance has to be made for possible sick leaves and other unexpected absences.

Table 4. Personnel requirements for different survey modes

Survey mode		Required personnel			
		Administration	Interviewers	Measurers	Data administration
Questionnaire	**Self-administration**	Yes	No	No	Yes
	Interview	Yes	Yes	No	Yes
Questionnaire and physical examinations		Yes	No/Yes	Yes	Yes

Table 4 lists the main personnel groups required for different survey modes. In a self-administered questionnaire survey no interviewers are needed and the survey can be managed with administrative personnel that take care of mailing of the questionnaires and reminders and personnel for data administration. When the questionnaire survey is conducted using interviews, the requirement of personnel increases. Now, interviewers are needed in addition to administrative and data administration personnel.

Surveys with physical examinations require more personnel than questionnaire surveys. In an examination survey administrative personnel are needed to arrange the appointments for

survey examinations, and possibly to mail out the questionnaires. Interviewers are needed if questionnaires are not self-administered. Specially trained personnel are required to take physical measurements and to handle collected samples. And just like in any other survey, data management personnel are required to take care of collected data.

Conclusions: Proper amount of personnel with required skills will decrease the workload of personnel, and through that improve the quality of the survey by minimizing the errors due to hurry and work related stress.

2.5.3.2 Selection of personnel

Several studies have shown that the characteristics of interviewers may affect the willingness to participate in the survey and to answers the questions. [25, 74, 75] Therefore, the selection of interviewers should be made with this in mind. More information about the interviewer effects is given in *Chapter 2.7.2.1*.

Generally, interviewers, especially for face-to-face interviews, should have clean and neat appearance and good manners, i.e. be friendly, direct, honest and interested in what the respondent says. [25, 74] With respect to age, gender, ethnicity, and social status the interviewer should have characteristics similar to the respondent [76]. This is especially important for sensitive questions [74].

As the interviewer is working as a mediator between the written questionnaire and the respondent, he/she should be able to read out questions clearly without any hesitation and without stumbling over words [25].

For surveys where different physical measurements are taken, the personnel have to include measurers. Good choices are nurses or medical students who are specially trained for the survey measurements. In a survey where blood specimens are taken, personnel are needed for handling the samples and possibly also for transferring samples from the survey site to the laboratory or storage. [17]

Although the personnel selection process is difficult and time consuming it should be taken seriously, since the success of the survey is heavily dependent on the performance of the survey personnel [35]. The personnel have to be motivated and have to show positive attitude even during difficult times of the survey. They also have to be willing to learn new things in a relatively short time.

For example, if personnel are not motivated, they will easily slip on the survey protocol, lowering the reliability and accuracy of the results.

Conclusions: Selection of the personnel will affect the final survey outcome.

2.6 Questionnaire design

During the preliminary stage of the survey planning, the objectives of the survey have been decided through research questions. Now these research questions should be transferred to actual questions in the survey questionnaire. There is extensive literature on the questionnaire design, which lists possible pitfalls of questionnaire design and gives a number of "how to do" guidelines [5, 20-22, 25, 26, 77-82]. Some of the key points of these guidelines are reviewed here and the literature should be consulted for more details and examples.

Conclusions: The questionnaire design will affect the response rate and the reliability and accuracy of the information obtained by the questions.

2.6.1 What to ask?

During the formulation of survey objectives, the number of research questions has been selected so that they will form a uniform set. Now the research questions are transferred to the actual questions.

Table 5. Defining required information to derive needed indicators

Indicator	Required information to derive indicators				
	Is the person hypertensive?	**Is the person using antihypertensive drugs?**	**Has the person been told by a health professional that he has elevated blood pressure?**	**Has the person had his blood pressure measured during the past year?**	**Blood pressure (systolic and diastolic) measurement results**
Prevalence of hypertension	Yes				Yes
Prevalence of antihypertensive drug treatment in population		Yes			
Prevalence of antihypertensive drug treatment among hypertensives	Yes	Yes			Yes
Awareness f elevated blood pressure			Yes		
Proportion of the population with blood pressure measurement in the past year				Yes	
Population mean (standard deviation) systolic/diastolic blood pressure					Yes

The first step is to identify what kind of information is needed to answer the research questions [25]. One way of doing this is to start from the indicators, which are needed when survey results are reported.

A simple cross-tabulation of indicators and required information can be used to define necessary information for the indicators. Sometimes one question/measurement may

contribute to several indicators. *Table 5* gives an example for the cross-tabulation of needed indicators and the information required deriving them.

When one knows what to ask, required information needs to be transformed into questions. Before starting to create new questions, already existing questions that are used in the other surveys (national and international), or provided by different guidelines should be reviewed to see if they are suitable for the survey.

Conclusions: Decision about what to ask and measure will affect the possibilities to answer the survey objectives.

2.6.2 How to ask?

There are occasions, when questions do not exist from previous surveys, or national/international survey guidelines, or where such examples just are not suitable for the survey. In such a situation new questions need to be developed. Questionnaire design is a combination of subject knowledge, psychology, linguistic know-how, and artistic vision just to mention a few of the needed skills.

Generally, a good questionnaire has been designed from the point of view of the respondents. A well-designed questionnaire minimizes the respondents' burden by being relatively fast to fill in. It also does not require too much thinking (retrieving information) on the part of the respondent and is not confusing or threatening. It respects the respondent as a person, not embarrassing him/her with too personal questions or with questions he/she does not understand. [25] On the other hand, the questionnaire also has to measure the intended outcome correctly and accurately.

Key elements of questionnaire design are language and wording of the questions, selection of response alternatives, formulation of sensitive questions, recall bias, order of individual questions, use of jump rules and the length of the questionnaire. If new questions are prepared

for a multicultural setting, the cultural adaptation of the questions needs to be taken into account [24].

Conclusions: The formulation of the questions will affect the willingness to respond and the accuracy of the answers.

2.6.2.1 Language and wording of the questions

The respondent starts his/her response process with comprehension of the question. The proper wording of the question is essential to ensure that the respondent has understood the question correctly. There are a few basic guidelines to help in the formulation of understandable questions.

The questions should be in the native language of respondent. This will increase both the response rate and the comprehension of the questions. When questions need to be translated from one language to the other, translated questions should correspond to the original question not so much in the wording but more in the intended meaning of the question [24, 81].

If the respondent cannot understand the meaning of the words used in the question he/she cannot make a decision about the answer to it. In such a case the respondent often omits the whole question and may not answer subsequent questions either. Thus, the words used in the questions should be familiar to the respondent. Professional terms or slang should be avoided unless the questionnaire is targeted for some specific group which is expected to know the terminology/slang [25, 74, 78, 79, 81]. Abbreviations should be left out of the questions, as they may not be familiar to some of the respondents [25, 81]. Words, which could be insulting to ethnic- or gender-related issues should be avoided, as well as words with strong positive or negative images [74].

> *Example.* An often used filtering question in smoking questionnaires is *"Have you ever smoked at least 100 cigarettes in your entire life?"*. A study conducted in New York City, USA examined how people actually interpreted this question

[83]. The question interpretation focused on two words “smoked” and “cigarettes”.

Among the respondents, 46% interpreted smoking meaning only inhaled puffs whereas 54% though it would mean all puffs, whether or not inhaled. Another 23% of respondents considered only finished cigarettes as smoked cigarettes, 23% counted any cigarette smoked completely or partly and 54% thought that a smoked cigarette is any smoked cigarette including also just one puff.

In most languages, there are words, which can be used interchangeably. In questionnaires, only one of them should be used consistently through out the questionnaire, otherwise the use of different words in different places may lead respondent to think that they mean different things [81].

Precise wording should be used in formulating the questions. If the question is vague, the respondent does not know how to respond. [74]

Example. Question *“How many people live in your household?”* does not specify if the respondent should include himself or not. A more specific wording would be *“Including yourself, how many people live in your household?”*. [74]

When preparing the questions, ask only one thing at a time [74]. Including several questions into one question will increase respondents’ burden and may cause confusion resulting in missing the answer to the entire question or some parts of it.

Example. Do not ask *“Have you ever been diagnosed to have heart disease or stroke?”*. Instead, spit the question into two; *“Have you ever been diagnosed to have heart disease?”* and *“Have you ever been diagnosed to have stroke?”*. [74]

Occasionally the questions may require additional background information or explanation to help the respondent with his/her answer. When formulating these additional instructions one has to avoid influencing the respondents’ answer. [74, 81]

Sometimes, it is desirable to emphasise some words in a question. This may be needed to make the question easier to understand. Using emphasis like bold, italic or capital letters or underlining is acceptable but one must be careful not to draw the respondent's attention to wrong issues. [74]

In general, the comprehension of questions is easier if they are short and simple. Sometimes, longer explanations are needed to assure the proper understanding of the questions. This is especially important if the questionnaire is self-administered.

Conclusions: The language and words used in the questions will affect the understanding of the questions, i.e. the accuracy of the answers and the response rate.

2.6.2.2 Response alternatives

Questionnaires can have questions with readily defined response alternatives (close-ended) or open questions where respondent can write down his/her own response (open-ended). Both alternatives will collect the needed information but in certain situations one works better than the other.

A close-ended question lists the response alternatives for that particular question. The response alternatives need to be carefully selected, as they can help the respondent to better understand the questions. If poorly selected, they might make answering more difficult. [74, 84]

For close-ended questions, the response alternatives can be categorized into three groups: responses on nominal, ordinal, or interval scale. Nominal scale response alternatives are just labelled responses without ordering or assigning any quantitative value to them (example: sex). Ordinal scale response alternatives can be ordered but no quantitative value can be assigned to them (example: education). Interval scale response alternatives can be labelled, ordered and also have standard units of measurement assigned to them (example: age group). [74]

In all cases, the response alternatives should cover all possible choices, but they should not be overlapping. At the same time, the list should not be too long and there should be clear difference between the response alternatives. If the list has many response alternatives the difference between them may become marginal, and then respondents may not be able to distinguish between them. [25, 74]

For interval scale response alternatives, there has to be clear quantitative difference between categories and each response category should be equally wide to avoid skewed response distribution. Response categories should also have a reasonable number of responses, i.e. response categories should be wide enough. [74]

Long lists of response alternatives are difficult, especially in telephone interviews, when the respondent cannot see the list and has to memorize the alternatives [25]. For telephone interviews the number of response alternatives should not exceed six. In questions intended for face-to-face interview or self-administration, the number of response alternatives may be as high as 15. In some special cases, the number of response alternatives can be up to 20 if show cards can be used to help the respondent to choose the answer. [74]

In general, nominal scale response alternatives are the most difficult for respondents, since the differentiation between answer categories is not always clear. Ordinal and interval scale response alternatives are more specific than nominal scale response alternatives and therefore they are easier to answer. [25]

The order of the response alternatives can affect the answers. It seems that in self-administered questionnaires, respondents are more likely to choose from among the first alternatives of the list whereas in interviews they are more likely to select among the later alternatives. Long lists of alternatives are more prone to this kind of behaviour. [25]

An open-ended question allows the respondent to write down everything he/she feels necessary for the answer. This will easily produce a lot of irrelevant and repetitious information but also makes it possible to gain more information than through the closed-

ended questions. Open-ended questions also make comparisons and coding of the answers more difficult. [74]

Open-ended questions are demanding for the respondents, who need to explain their attitudes or behaviours. This requires a certain amount of communication skills and much more time than answering to the close-ended questions. These demands on the respondent may result in higher refusal rates. Open-ended questions also provide for a wide range of answers, which can be difficult to interpret and compare. Data entry for open-ended questions is more difficult and time consuming than for close-ended questions, since each response has to be interpreted before coding. Data entering personnel need detailed instructions for coding each answer. The possibility of coder bias is much higher than with close-ended questions. [25, 74]

Occasionally, combinations of closed-ended and open-ended questions are used. These questions have a list of alternatives one of which is *"other, please specify"*. The respondent can then give an answer that is not already listed. Most of the respondents choose one of the already listed answer alternatives. Again, coding of this kind of question requires detailed instructions for the open-ended part. Also, the comparison and analysis of this kind of information is difficult. [25]

Regardless of the type of answer format; close-ended, open-ended or combination of these two, the respondents should not need to make any difficult calculations, like calculating percentages, in order to reach their answer. Always, only totals are asked, which the respondent can easily obtain by summing up. Percentages are calculated later from given totals by the survey organizers. [25]

The choice between close-ended or open-ended questions is not trivial. *Table 6* summarizes the advantages and disadvantages of the close-ended and open-ended questions. The close-ended questions are generally easier for respondent, i.e. reduce the respondents' burden, and they are also easier to code and compare during the analysis stage. Also, the given response alternatives, if properly selected, can help the respondent to remember issues, which he/she would not otherwise recall. For sensitive issues like income, the close-ended questions tend to work better. Instead of having to provide the specific amount of income the respondent can

select one of the alternative ranges. The open-ended questions can detect moderate differences between the respondents but at the same times, they require high communication skills, which may sometimes reduce the response rate.

Table 6. Advantages (+) and disadvantages (-) of close-ended and open-ended questions

Features	Type of response alternatives	
	Close-ended	**Open-ended**
Easy to answer	+	-
All response alternatives can be obtained	-	+
Sensitive questions	+	-
Recall of answers	+	-
Higher response rate	+	-
Responses easy to code	+	-
Requires detailed coding instructions	-	+
Responses easy to compare	+	-

Conclusions: Response alternatives will affect the accuracy of the answers and possibly the item non-response.

2.6.2.3 Sensitive questions

Sensitive questions cover socially sensitive issues or issues individuals may consider as private information. If questionnaire contain such questions they may cause difficulties as they easily get the respondents guarded and/or make them feel uncomfortable. Respondent may refuse to answer these questions or reject even the entire questionnaire.

Sensitive questions tend to work better in the self-administered questionnaires than in interviews. The self-administration ensures the confidentiality of the results, which is missing from the interview. Also, using broad answer categories, instead of asking for specific details works better for sensitive questions [25].

Generally, sensitive questions should be placed towards the end of questionnaire. This will ensure that in case of negative reaction to a sensitive question, i.e. stopping to fill in the questionnaire, the information from the beginning of questionnaire will still be available. For interview, placing the sensitive questions towards the end may also help. If the respondent has formed a positive, trusting relationship with the interviewer during the earlier part of the interview, he/she may be more willing to answer also the sensitive questions later on. [25, 74, 82, 84]

Conclusions: Sensitive questions may lower the response rate.

2.6.2.4 Recall bias

Human memory records events as they occur and sometimes their recollection is easier than at other times. Recent events are generally easier to recall than things that happened a long time ago (*Figure 9*). On the other hand, if a person has dementia, he/she may forget what has happened a few minutes ago but remember clearly things from decades ago. The accuracy of recall is often related to the length of the recall period. The longer the recall period the worse is the accuracy of recollection. [82]

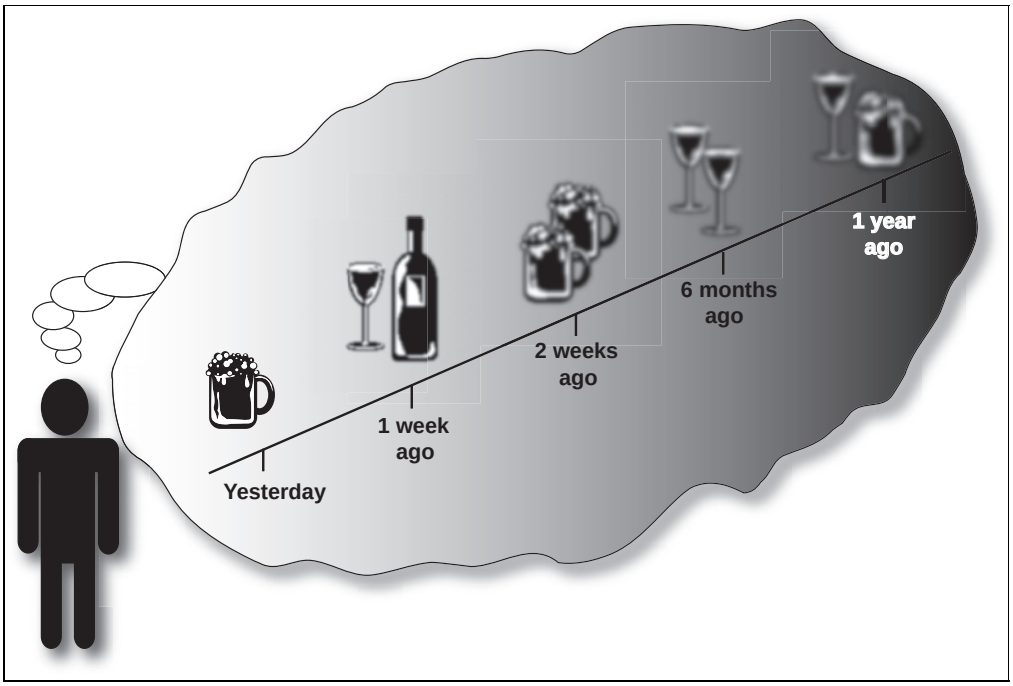

Figure 9. Recall bias

The longer the recall period the higher is the risk that the respondent projects events into the time frame of the question. This phenomenon is called telescoping effect. [82, 84]

The recall of the events can be assisted by adding aids to the questionnaire and by ordering of the questions. For example, questions about the events can be linked to holidays or national festivals, or the respondents can be asked to use the calendar to set events in the correct time frame. [82, 84] When questions about past events go back step-by-step, starting from the most recent ones, it may help the respondent to remember events further in the past.

Conclusions: Too long recall period may create bias.

2.6.2.5 Order of questions

The order of individual questions in the questionnaire is important. There is evidence, that the question order has an effect on response rate [82, 84] and the willingness to answer sensitive questions [74, 82, 84].

Often, the first question defines the respondents' attitude towards the whole questionnaire. Therefore, the choice of the first questions is very important. The initial questions should be the close-ended, because this type is easier to answer than the open-ended questions. The first question should also be easy to understand and should not deal with an embarrassing subject, and, if possible, should not ask for personal opinions. [25, 74] Complicated and sensitive questions at the beginning may lead the respondent to the decision not to answer any of the questions in the questionnaire. [74, 82, 84]

Starting with easy questions and gradually moving toward ones that are more difficult to answer will help to collect data at least for some of the questions. When difficult questions are at the end of questionnaire and if the respondent does not want to answer them, at least some data for earlier questions have been collected. In interview questionnaires, the respondent may establish a trusting relationship with the interviewer during the early easy questions and then be more willing to answer also more difficult questions towards the end of the questionnaire. [82, 84]

Questions should be grouped by subject area. Within a subject area; questions with the same nature (yes and no questions, long lists, etc.) should be kept grouped together. This helps the respondents with their answers. [25, 74] In grouping questions, it is important to remember that too many similar questions in a row may cause reflexive responses. [74]

A properly organized questionnaire can also reduce the respondents' burden by filtering out questions that are not relevant for some respondent [81, 84]. In questionnaires, there are often sections, which should be filled in only if certain criteria are met. For example, detailed information about smoking habits should only be asked from smokers. The questions should be organized so that there is first a filtering question, which ask about the respondents'

smoking status. This can then be followed by more detailed questions about the smoking habits. Respondents who do not smoke can be instructed to skip the detailed smoking questions and move to the next group of questions.

Example. Smoking questionnaire with the filtering questions [17].

SMK1 **Have you smoked at least 100 cigarettes, cigars or pipefuls in your lifetime?**

1 = Yes

2 = No (*Go to question SMK8*)

3 = Uncertain (*Go to question SMK8*)

SMK2 **Have you ever smoked daily (= almost every day for at least one year)?**

1 = Yes

2 = No (*Go to question SMK8*)

3 = Uncertain (*Go to question SMK8*)

SMK3 **Do you now smoke?**

1 = Yes, daily *(Go to question SMK5)*

2 = Yes, occasionally

3 = Not at all

SMK4 **When did you stop smoking daily? (If you have quit smoking several times, give the time when you last stopped smoking daily?)**

1 = Today or yesterday

2 = 2 days - 6 days ago

3 = 1 week - less than 1 month ago

4 = 1 month - less than 1 year ago

5 = 1 - 5 years ago

6 = More than 5 years ago

Go to question SMK7

SMK5 **On average, how many times do you smoke per day (= number of cigarettes, cigars, pipefuls of tobacco etc.)?**

SMK6 **Which of the products do you <u>frequently</u> smoke?**

Manufactured cigarettes (1=yes, 2=no)

Self-rolled cigarettes (1=yes, 2=no)

Pipe (1=yes, 2=no)

Cigars (1=yes, 2=no)

SMK7 **Have you during the past year (12 months) been advised by a health professional to stop smoking?**

1 = Yes

2 = No

3 = I have not smoked during the past 12 months

SMK8 **Are you exposed to indoor tobacco smoke at home?**

1 = Yes

2 = No

SMK9 **About how many hours per day are you exposed to indoor tobacco smoke at your workplace?**

1 = I do not work outside the home

2 = Almost never

3 = Less than one hour a day

4 = 1-5 hours a day

5 = More than 5 hours a day

When moving from one subject group to the other, there should be clear indication that the subject group has changed. The information about the transition can be indicated as sub-heading in the self-administered questionnaire or by the short phrase *"Next I would like to ask*

you about…" in the interview. This brief information helps the respondent to get orientated to the new of subject area. [25]

Conclusions: The order of questions may affect the recall of events and the respondents' burden through filtering questions. The order of questions may also affect the response rate in case of sensitive questions.

2.6.2.6 Jump rules

Jump rules (skip-patterns) are used to avoid people answering questions irrelevant to them [74]. For example, someone who has never smoked does not need to answer any questions related to present or past smoking habits. The formulation and presentation of the jump rules in the questionnaire may have an effect on the respondent's actions. If jump rules are not clear and unambiguous, the respondent may end up filling in questions not meant for him/her or jump over the questions he/she should fill in.

When the questionnaire is filled during an interview, the jump rules should be clear to the interviewer, so that he/she does not ask unnecessary questions from the respondent. If the interview is done using computers for data entry (computer aided personal interview = CAPI or computer aided telephone interview = CATI), the jump rules should be programmed into the system so that the questions, which should not be asked, are not even shown to the interviewer.

When the questionnaire is self-administered, the jump rules need to be part of the questionnaire. Often, written instructions are not enough for all respondents to follow them correctly (*Figure 10*). Different visual aids, like arrows and question layout can be used to guide the respondent from question to question [25].

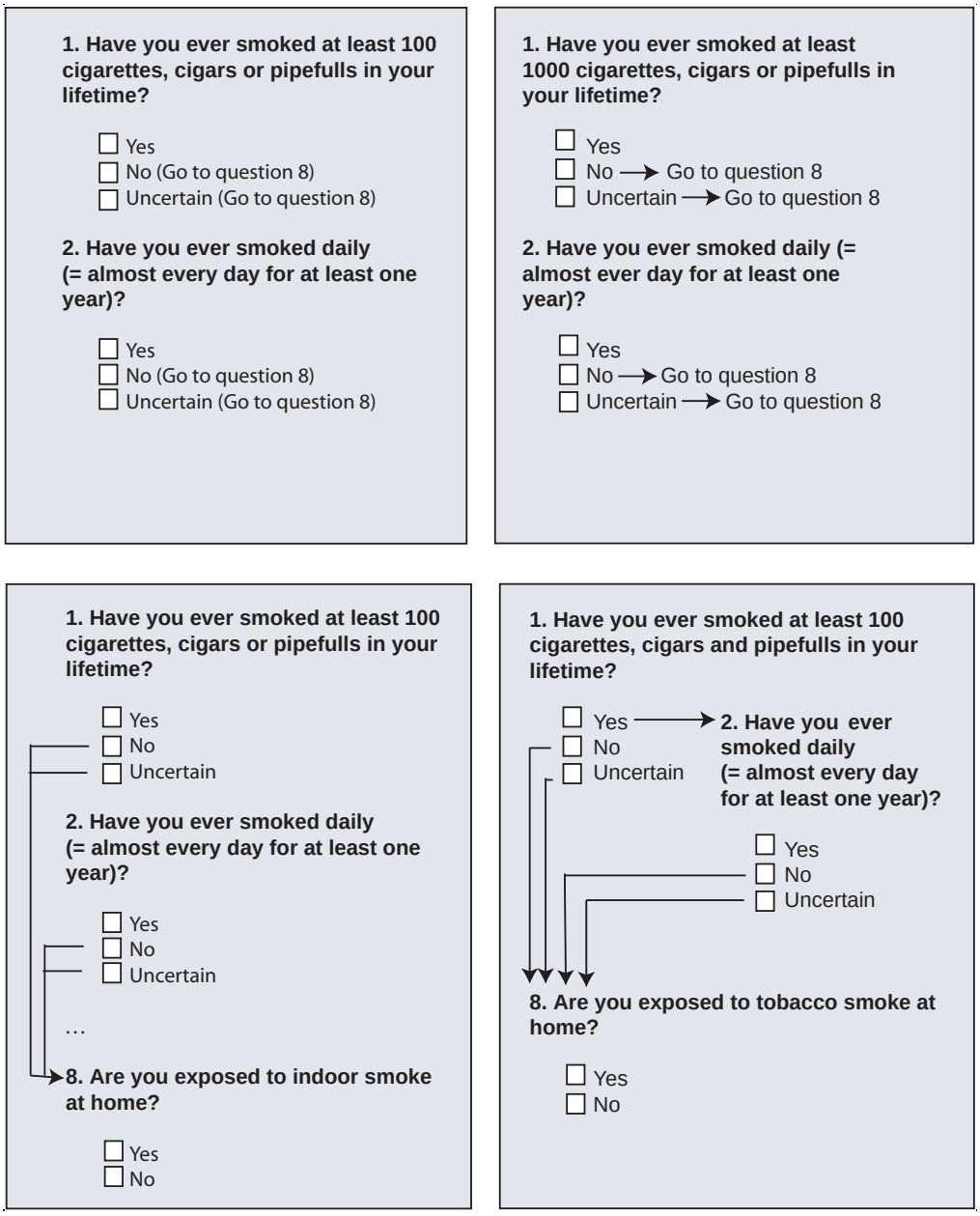

Figure 10. Different alternatives for the visual aids of the jump rules

Conclusions: Properly used jump rules reduce the respondents' burden by omitting the questions not relevant for some respondent groups.

2.6.2.7 Length of the questionnaire

The length of the questionnaire is directly reflected in the respondents' burden and therefore in the response rate and the reliability of the responses (*Figure 11*). Finding the balance between the number of questions, response rate, and the reliability of responses is difficult.

Shorter questionnaires will increase the response rates and the quality of data but may lack the important questions for the determination of needed indicators [84]. On the other hand, long questionnaires tend to get lower response rates because respondents become reluctant to complete them or complete them only partially [85, 86]. Long questionnaires may also be tiring for the respondent. Even if they fill them completely, respondents easily get careless in their answers towards the end of the questionnaire and the reliability of the answers decreases [84, 87].

Nowadays, people get bombarded with all kinds of surveys and they are getting more selective about the use of their time to answering survey questionnaires. Time is considered valuable and it is easier to convince the respondent to spend time to answer fifty questions than 200 questions.

In general, the questionnaire length should be designed so that completing it should take maximally 20 minutes for telephone interview or 30 by self-administration or 45 minutes to one hour by face-to-face interview. Ideally, the telephone interview should takes only 10 minutes, the self-administration of the questionnaire only 15 minutes and the face-to-face interview only 30 minutes. [74]

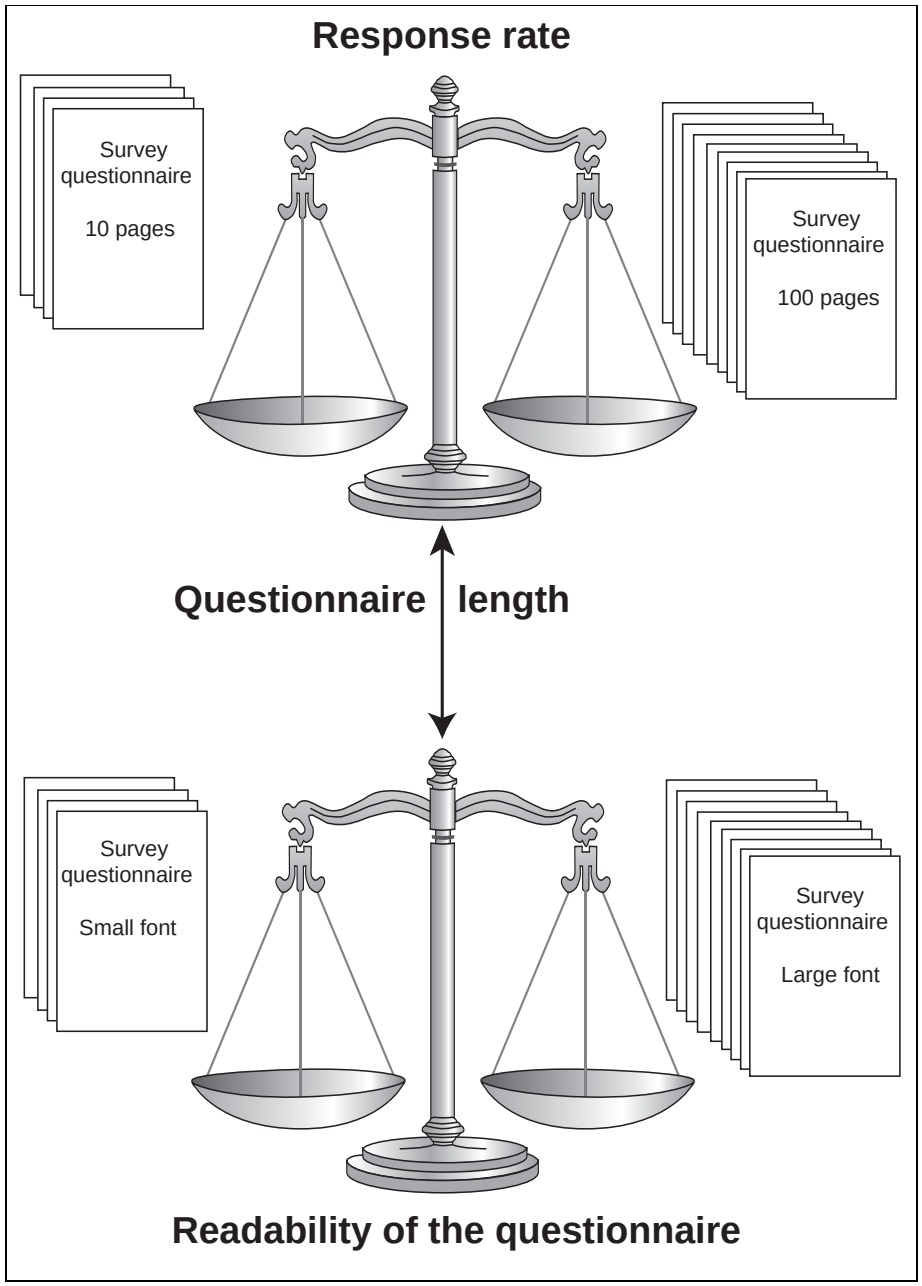

Figure 11. Balancing between the questionnaire length, the response rate, and the reliability of the responses

Conclusions: The length of the questionnaire will affect the response rate and the reliability of the responses.

2.6.2.8 Pre-testing, reliability and validity of the questionnaire

When new questions are prepared, they should be pre-tested and validated for the study population. Also, if the questions from other studies are adapted to the survey, they should go through the same pre-testing and validation as new questions. Questions, which provide accurate and reliable information in one population do not necessarily work in other populations due to cultural difference [24].

The purpose of pre-testing of the questions is to test the clarity, comprehensiveness, and acceptability of the questions. When testing the clarity of the questions, one tests whether the respondents understand the question correctly. When testing the comprehensiveness of the questions, one tests whether the respondents know all the words/terms used in the questions, that all required response alternatives are listed, and that the recall period is feasible. When testing the acceptability of the questions, one tests whether the questions are ethically and morally approved, i.e. whether questions are not too sensitive, do not offend the respondents' privacy, or cause too much burden for the respondent. [5, 25, 74]

The pre-testing of the questions should start with a critical review by the researcher. During this critical review, the researcher him/herself should fill in the questionnaire to see if he/she can fill it in without any problems. After that, a number of colleagues and friends should be asked to do the same and report back all the problems they encountered. [20]

The actual pre-testing is done using a small sample of 25 to 75 persons, representing the target population [74, 78, 79]. There are a number of methods, which can be used for pre-testing. Pre-testing can be conducted, for example, using cognitive interview [24, 26], behavioural coding [88], special probing [88], expert panel [89], or by conventional way [89], where experienced interviewers conduct a small number of interviews using the questions to be tested and they report their experiences and observations.

The reliability of the questions refers to the consistency of the questions, i.e. the questions should provide the same answer from the same person in repeated trials. To test the reliability of the questions, the same questions are asked from same persons more than once over time. [90, 91]

The validity of the questions refers to the correctness of the obtained outcome. The validity of the questions can be assessed against other existing data sources. Depending on the nature of the question, these data sources may be the administrative and other registers, other similar questions, etc. For some outcomes, there exist "gold standards", measurements, which are used to obtain the true values. The basic idea of the validation of the questions is to obtain data by the new questions and compare them with the data from another source provided by the same person.

It should be noted that the question could be reliable without being valid. Questions can produce the same answer when asked over and over again, but not measure the correct outcome.

> *Example.* It is of interest to measure the prevalence of hormone replacement therapy in the population (presently). The question used is *"Have you ever used hormone replacement therapy?"*. When the question is asked from the same group of women on two occasions, two weeks apart, they provide the same answer i.e. the question is reliable. When the questionnaire is validated against the medical records, the prevalence based on the question is higher than the prevalence based on medical records, i.e. the question is not valid. This discrepancy between the information provided by the question and the medical records is due to the wrongly specified time frame in the question. The question asks if a person has ever used hormone replacement therapy, but the interest is in the present use, i.e. the question is reliable but not valid.

Conclusions: Pre-testing, and testing of reliability and validity will increase the accuracy and reliability of the results.

2.6.3 Layout of the questionnaire

Layout of the questionnaire is as important as the questions. This importance is emphasised when the questionnaire is self-administered. If the questionnaire is difficult to read and the first impression is unprofessional, the respondent easily decides not to answer it.

Conclusions: Layout of the questionnaire will affect the response rate.

2.6.3.1 Paper questionnaire

The readability of the questionnaire can be improved by following the guidelines given for printed material that is prepared for visually impaired people. These guidelines list several basic issues like selection of typeface, type size, type weight, use of italic or capital letters, readability of numbers, selection of interlinear space, line length, line endings, word spacing, and paragraphs [92, 93].

The used typeface should be easy to read with clear difference between the different letters and numbers. Recommended typefaces are a sans serif (example: Arial or Helvetica) and a seriffed (example: Times or Century). The type size should be big enough for easy reading, i.e. 12 pt for usual reader and 14 pt or 16 pt for the visually impaired ones (*Figure 12*). Previously, bold text was used to make text more readable, as it increases the contrast between the text and the background paper. Now, the recommendation is to use medium weight or semi-bold and to avoid light weighted typefaces. [92]

Arial 12 pt	Times New Roman 12 pt
Arial 14 pt	Times New Roman 14 pt
Arial 18 pt	Times New Roman 18 pt
Arial 12 pt bold	**Times New Roman 12 pt bold**
Arial 14 pt bold	**Times New Roman 14 pt bold**
Arial 18 pt bold	**Times New Roman 18 pt bold**

Figure 12. Recommended typefaces and type sizes

For working-age adult populations, the type size 12 pt should be sufficient. When the questionnaire is targeted for older people (60 years +), the type size should be increased at least to 14 pt, preferably to 16 pt.

Italic type can be used to emphasise important words, but it should be used sparingly. The same applies to the use of capital letters. (*Figure 13*) Text written in capital letters is more difficult to read than text in normal-case. [20, 92, 93]

Have you ever smoked daily (= almost every day for at least one year)?

Have you ever smoked daily (= almost every day for at least one year)?

HAVE YOU EVER SMOKED DAILY (= ALMOST EVERY DAY FOR AT LEAST ONE YEAR)?

Figure 13. Effect of italic type and capital letters to the readability of the text

The space between lines should be big enough, so that lines are easy to follow. The Royal National Institute for the Blind (RNIB) recommends using at least +2 pt between lines for the type size 14 pt to 20 pt, while other recommendations instruct to add 25% to 30% of the type

size between lines (+3.5 pt to + 4.2 pt for the type size 14 pt). Generally, the longer the lines are the more space is needed between the lines to ensure readability. Ideally, the lines are maximally 50-65 (60-70 by RNIB) characters long and the words are not split between the lines, i.e. no hyphenation. [92, 93] Shorter lines make it less likely that the respondents skip words. [20]

Even spacing between words and clear difference between the lines and paragraphs increases readability of the text. The space between the paragraphs should be clearly larger than the space between the lines. [20, 92]

If the text is placed in columns, the space between the columns should be generous or in case this is not possible, there should be a vertical rule between the columns [92]. Visual aids, like pictures and arrows guiding the jump rules should be clear and located so that there is no place for confusion to which question they refer [74]. Questions should be placed so that they do not break between the columns or pages [74].

To improve the readability, the question itself should be in bold and the response alternatives in light font to separate them from each other (*Figure 14*). Also, when additional instructions for the question are needed, their font type should be different from the actual question. [20]

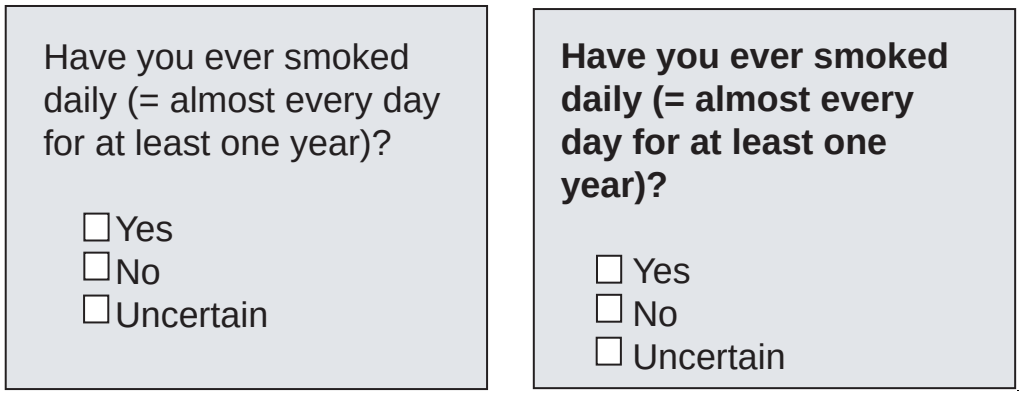

Figure 14. Readability of the question and response alternatives

Response alternatives should be listed vertically not horizontally, when ever possibly. [20]

Paper on which the questionnaire is printed and the colour of the print also affects the readability of the text. The general guideline is that the bigger the contrast between the background and the text, the easier is it to read. Therefore, the best combination is to use white paper and black text. The paper should be uncoated paper, because glossy paper will create glare, which makes it difficult to read. Paper should also be dense enough that the printed text does not gleam through the page (weight over 90 GMS). [92]

The questionnaire should always be bound, i.e. not have loose pages, which can get lost [74]. The questionnaire format should be similar to that of a conventional book or newspapers, because they are familiar to the people and therefore easy to read. If some unconventional format or folding is used, the respondent has to spend extra time and effort to figure out in which order to proceed. In western cultures, the most common format would be a vertical book or booklet. For other cultures, different formats may be more suitable. [20]

The front cover of the questionnaire plays an important role, since the decision to read the questionnaire is made based on the front cover and general appearance of the questionnaire. The front cover of the questionnaire should be attractive and have a professional look and not that of an advertisement brochure. [25, 74] The information on the front cover should be limited to the most essential information. An informative title, a graphical design or illustration (for example: organization logo), which can be linked to the study, the name(s) of the study's sponsor(s). The makeup of the cover page should also include the address where to return the questionnaire as well as the phone number where to get more information. [25] A covering letter sent with the questionnaire should be on a separate piece of paper. Listing the above information only in the covering letter is not sufficient, because the covering letter can easily get lost.

The back cover of the questionnaire also has an important role to play. It should thank the respondent for participation and invite comments with enough space provided for that. [25, 74]

The role of the questionnaire layout is especially important when the questionnaire targets older people, many of whom will have impaired vision.

The researcher has to balance between the readability and length of the questionnaire (*Figure 15*), both of which affect the response rate. If type size is made smaller the same number of questions can be fit on fewer pages but at the same time the readability of the questionnaire may be reduced. For some respondents this may be a reason for not answering the questionnaire. On the other hand, keeping the type size big enough to ensure good readability will increase the number of pages of the questionnaire, making the questionnaire appear longer. Again, for some respondents this may be a reason for not answering the questionnaire.

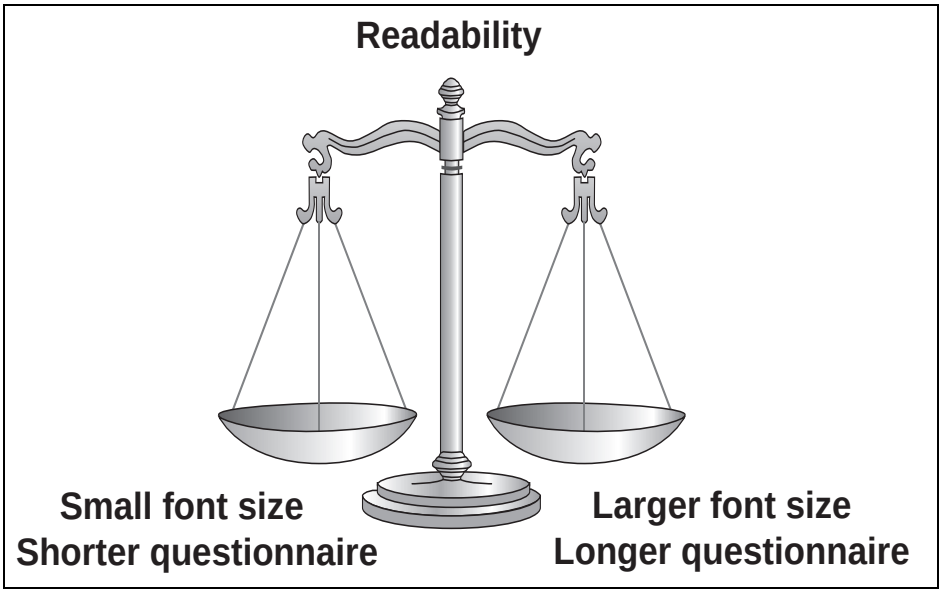

Figure 15. Balance between the readability, the font size, and the questionnaire length

2.6.3.2 Electronic questionnaire

Electronic questionnaires are usually questionnaires on the Internet. When they are used, one has to remember that in addition to the respondent there is the second element, the computer and the software within the computer, which needs to be taken into consideration while

preparing the questionnaire. The electronic questionnaires should be usable regardless of the speed of the Internet connection and used software. This requirement sets some limitations for the design of the electronic questionnaires, since HTML [94] is the only common language automatically used by all web browsers. If questionnaires are designed to use for example Java [95], the web browser settings have to allow the use of Java and sometimes additional software/plug-ins have to be installed before the questionnaire can be viewed and filled-in.

Each page of the questionnaire loaded at once should be small enough to work well even with slow Internet connections. If respondents have to wait long for each page to load, they easily move to other things and will not answer the questionnaire at all. [20]

The layout of the questions should be designed so that the screen size does not affect the line breaks, or cause misaligned text. Also, the text should be designed to fit on a page without requiring horizontal scrolling. [20]

The text and background colours should not be fixed, but left for the respondent to decide by the settings of their web browser. Also, the size of the text should not be fixed. [96]

2.7 Questionnaire administration

There are two main administration modes for the questionnaires: self-administration and interview. When the questionnaire is self-administered, the respondent him/herself fills either a paper or electronic form. During the interview, which can be conducted either face-to-face or through telephone, the interviewer reads the questions and codes the answers given by the respondent. The interviewer's questionnaire for coding the answers can be either in paper or electronic format.

The choice of administration mode may lead to the under-coverage of the target population. In a population with low telephone coverage, choosing the telephone interview as administration mode will limit the population coverage to those with telephones leaving the majority of the population out.

Conclusions: The questionnaire administration mode may affect the response rate and the accuracy and reliability of the responses. Also, the selection of the questionnaire administration mode may affect the representativeness of the population.

2.7.1 Self-administration

Self-administered questionnaires are usually mailed to the survey participants. Questionnaires can also be delivered personally (drop-off surveys), or handed out during the survey examinations to be filled in on-site or later at home and mailed back. Modern technology has introduced a new way for self-administration of the questionnaires. Nowadays, the questionnaires can be sent to the people not only through regular mail but also by e-mail, and the self-administered questionnaires no longer have to be paper forms, they can also be electronic forms, which are filled in on the Internet (web surveys).

Self-administration of the questionnaire requires a high literacy level of the population to obtain acceptable response rate. Whenever the literacy level is low or the level of understanding the text is low, the questionnaire should be filled in by interview.

Mail surveys are a good choice when the target population is widely distributed, when there are only limited financial resources or number of personnel to conduct the survey, and when fast turnaround time is not required. For mail surveys the questions should also be mainly close-ended and the questionnaire should be relatively short. Questions requiring privacy for answering favour mail surveys. [97]

The biggest limitation for the web surveys is sampling. Web surveys cannot be conducted in a representative population sample of the general population, as in many populations, the major part of the general population does not have the Internet access. The sampling problem exists even when the survey would be aimed at Internet users, because there are no comprehensive sampling frames to use. Also, the technical requirements can limit the use of web surveys,

since different standards between operating systems, web browsers and servers will require extra efforts to get the web forms to work the same way for everyone. [20, 98]

Based on the statistics by the International Telecommunication Union [99], the availability of PCs and the use of the Internet vary considerably between countries (*Table 7*). These statistics will not provide the information about the household level coverage of the Internet connections but will provide the population level information. For example, at the population level, the Internet use varies from 1% to 67% in Europe.

Table 7. The ownership of PCs and the use of the Internet in 2003 (variation in brackets)

Continent	PCs per 100 inhabitants	Internet users per 10.000 inhabitants
Europe	21 (1-71)	2373 (98-6747)
Africa	1 (0-16)	148 (10-1452)
Americas	29 (1-66)	2593 (96-5514)
Asia	4 (0-62)	674 (6-6034)
Oceania	42 (1-56)	3764 (50-5262)
World	10	1107

For example, in the USA in 2000, 51% of the households had a computer and 42% of the households had an Internet connection at home. Both, the ownership of computers and Internet access vary between age groups (most 25-44 years olds, least 65+ years olds), by race (most Asians and Pacific Islanders, least blacks and Hispanics), by educational level (most among those with bachelors degree or more, least among those with less than high school), and by income. [100] In 2002, 77% of Danes, 63% of Finns, 78% of Norwegians, and 75% of Swedes had a computer at home. At the same time, 61% of Danes, 53% of Finns, 73% of Icelanders, 61% of Norwegians, and 68% of Swedes had Internet connection at home. In all Nordic countries, these proportions were higher among the younger age groups (19-29 years), men, and individuals with higher education. [101]

Other disadvantages of web surveys are the technical complexities, page design effects, greater chance of getting partial responses unless the web form is programmed not to accept partial replies, and tenuous open-ended responses [20, 98].

On the other hand, the advantages of web surveys are their low cost after the establishment of the system, quick turnaround time, and collapsed geographical boundaries. Using the Java language and other Internet tools constraints to the response alternatives and jump rules can be programmed into the questionnaire, reducing data entry errors and ensuring higher quality. [98]

The elderly, persons from lower income groups, and women tend to prefer the traditional paper form instead of a web form [102, 103]. This is most likely due to the level of computer knowledge in these population groups.

There is not yet solid evidence about the effect of web surveys on response rates. Some studies have shown higher response rates for web surveys than for traditional mail surveys, but other studies have demonstrated opposite results. [98, 103, 104] These comparisons are usually done within a business or administrative office, where a complete sampling frame does exists, and every one has an equal opportunity to use computers and the Internet. It looks like the biggest reason for not completing the web forms is the technical difficulty [104] and after that the concern about the privacy [105].

2.7.2 Interview

Interviews are usually conducted either face-to-face or by telephone using paper forms or computers with electronic forms to code the answers provided by the respondent. When computers are used for data collection, one is talking about the computer aided personal interviews (CAPI) and computer aided telephone interviews (CATI).

Face-to-face interviews tend to have higher contact rate than the telephone interviews, since in many countries obtaining the correct phone number is more difficult than obtaining the

correct address [106]. Also, the coverage of telephones may be low, especially in some population groups [107]. Another problem with the telephone interview is the answering machine, which picks up the calls. Owners of telephone answering machines are usually young, unmarried with high socio-economic status [108]. The telephone interviews also eliminate deaf people. For example, in the USA, approximately 2% of the population cannot hear normal speech. Most of these people do not have telephone. [109]

The response rate of face-to-face interviews is generally higher than of telephone interviews [106, 110]. It is more difficult to say no when facing an interviewer in person rather than hanging up the telephone.

The identification of the respondent is easy in face-to-face interviews because the interviewer can ask for identification before starting the interview [25]. In a telephone interview, the interviewer has to accept that the person on the line identified him/herself correctly. Telephone interviews tend to work better for sensitive questions [110].

In telephone interview, questions with long lists of answer categories should be avoided since their understanding and processing by the respondent may be difficult (no visual aids). This applies also to questions that depend on any visual aids like maps, and diagrams. [25] Also, open-ended questions tend to be difficult in telephone interviews, resulting in truncated answers in comparison with face-to-face interviews [110].

Face-to-face surveys are much more expensive than are telephone interviews. [111] The interviewer has to travel from house to house to conduct the interviews. During the time when the interviewer moves from one place to the other, the telephone interviewer can easily conduct one or two additional interviews.

2.7.2.1 Interviewer effect

Selection and training of the interviewers is important since the interviewers can affect the answers given by the respondents. This influence is most often based on the general appearance of the interviewers and the ways in which they read the questions and probes.

The effect of the interviewer characteristics like age, sex, race, educational level, social class, and ethnic or religious background on the respondents' behaviour has been studied. It appears that religion or ethnicity of the interviewer does not generally have an effect on the answers unless the questions are related to religion/ethnicity. Also, the race of the interviewer has been found to have a similar effect as religion and ethnicity. There is no clear evidence that the social class or educational level of the interviewer would have an effect on the answers given by the respondents. [75]

As rule of thumb would be that the interviewers in their basic characteristics should be as close to the respondents as possible. Similar background of interviewers and respondents helps to minimize the interviewer effect. [76]

If interviewers are not trained properly, they can cause bias in the results by not reading questions as they are written, using wrong probing or adding their own comments to the questions, or just by miscoding the answers given by the respondent [75]. Often, interviewers tend to read too fast when asking the questions [112]

2.7.3 Pros and cons of interview and self-administration

It is not clear cut whether self-administration and interview is better, since both modes have their advantages and disadvantages. The final decision, which one to use has to be based on the evaluation of pros and cons in each situation separately. *Table 8* summarizes the characteristics of self-administration (mail and web) and interviews (telephone and face-to-face).

When the target population is spread over a wide geographical area, mail [19] and web surveys work best. Also, telephone surveys are good choice for covering large geographical areas.

If fast replies are expected, the telephone and web surveys are the best alternatives. The mail and face-to-face surveys are often equally fast. However, if the face-to-face survey is conducted in the geographically widely dispersed population, the mail survey may be faster. [19]

For the mail and face-to-face surveys, there usually are good sampling frames covering the intended target population. For web surveys, the sampling frames are so far missing and for telephone surveys, there often are coverage problems in the sampling frames. Due to the availability of sampling frames, the mail and face-to-face surveys often are well representative for the general populations. The telephone survey representativeness for the general population is dependent on the coverage of the sampling frame. For the time being, the web surveys are not representative for the general population.

Both mail and web surveys require a high literacy level, since the respondent has to read the questions. For telephone and face-to-face surveys, there are no specific requirements about the literacy level as the interviewer reads the questions. Mail surveys also require a well functioning postal system.

When the questionnaire is self-administered, the respondent decides the time and place as well as the order and pace of questionnaire completion. In an interview, the interviewer determines the order and pace of the questionnaire completion and often also the time and place of interview. [25] The sequence of questions can not be controlled in mail survey, where respondent can jump from one question to other as they please. [19]

Table 8. Characteristics of the mail, web, telephone, and face-to-face surveys

Characteristics	Self-administration		Interview	
	Mail	**Web**	**Telephone**	**Face-to-face**
Geographical spread of the target population	Wide	Wide	Wide	Narrow
Speed of data collection	Slow	Fast	Fast	Slow
Availability of the sampling frame	Good	Poor	Varies	Good
Representativeness for the general population	Good	Poor	Varies	Good
Literacy level requirement	High	High	None	None
Working postal system	Yes	No	No	No
High telephone coverage	No	No	Yes	No
High Internet coverage	No	Yes	No	No
Respondent chooses the time and place that is best for him/her to fill in the questionnaire	Yes	Yes	No	No
Control of the question sequence	No	Yes	Yes	No
Identification of the respondent	No	No	Some	Yes
Response rate	Low	Low	Medium	High
Item non-response	Yes	No	No	No
Long questionnaire	Yes	Yes	No	Yes
Open-ended questions	Yes	Yes	Yes/No	Yes/No
Visual aids for the questions	Yes	Yes	No	Yes
Socially desirable answers	No	No	Some	Yes
Privacy for the sensitive questions	Yes	Yes	No	No
Long list of the answer alternatives	Yes	Yes	No	Yes
Jump rules	Difficult	Yes	Yes	Yes
Questionnaire layout	Yes	Yes	Some	Yes
Possibility for probing and other clarifications if needed	No	No	Yes	Yes
Interviewer effect	No	No	Yes	Yes
Language versions	Difficult	Yes	Yes	Yes
Personnel requirements	Low	Low	Medium	High
Cost	Low	Medium	Medium	High

The identification of the respondent is easiest during the face-to-face surveys and to some extend is also possible during the telephone survey. In the mail and web surveys, one really does not know who actually filled in the questionnaire. [19]

The response rate is usually highest in the face-to-face survey, followed by the telephone survey, the mail survey and is lowest in the web surveys [19, 113-117]. The item non-response is higher in the mail surveys than in the telephone or the face-to-face surveys, since during the interview, the interviewer can control the item non-response [19, 117].

In order to achieve high response rate the questionnaires to be used in mail and telephone surveys should be simple and shorter than the ones used in the face-to-face interviews. Open-ended questions tend to be difficult in telephone and face-to-face interviews but work better in mail surveys, as long as they are simple. On the other hand, some people have difficulty writing their answers to the open-ended questions even though they can tell them orally. [19]

In a face-to-face survey, all kinds of visual aids can be used to help the respondent. This gives many more alternatives for the questions than in a telephone survey, where all the questions are verbal and have to be understood without any visual aids. The mail surveys, in this respect, are very much like the face-to-face surveys. [19]

The answers obtained by self-administration and interview are quite comparable, but the interviews tend to give more socially desired answers than the self-administration [19, 76]. This effect is stronger in women, among younger adults and when there is smaller age difference between the respondent and the interviewer. [76]

Sensitive questions tend to be easier to answer by self-administration than through interview because self-administration gives more privacy. [25, 84]

Older respondents have more difficulty answering questions over the telephone than through self-administered questionnaires. Older people also tend to use last alternative more often than younger people (order effect). People aged 75 years and older have significantly more first and last answers in the telephone survey than in the mail survey. In the telephone

surveys, the respondent may select the easy way out by choosing the first acceptable answer, agreeing with an assertion, or selecting a status quo answer. [118] The telephone interviews tend to give more *"no"* answers to simple *"Yes/No"* questions than self-administered questionnaires. [119]

The accuracy of responses to questions about physician visits is lower in telephone surveys than in mail surveys, having more underreporting. In mail surveys, underreporting increases by age, is higher among men, among the higher socio-economic class, in smokers and among those who do not have any chronic conditions and who are not using regular medications. In telephone surveys, underreporting is highest among young (17-35 years), men, lower socio-economic class, smokers, and people without any chronic conditions or regular medications. [116]

Jump rules (screening questions) are most difficult in mail questionnaires where the respondent has to make the decision to skip some questions and to move to the next section. In interviews, the interviewer will make this decision for the respondents [19] and in web surveys, jump rules can be programmed directly into the web form.

Effects of questionnaire construction (order of questions, lead-in statements, subtitles, etc.) are least for face-to-face interview, since the interviewer can provide additional information if needed. In a telephone survey, the questionnaire construction is more important than in the face-to-face survey because all communication is verbal. Questionnaire construction is most important in mail and web surveys. [19]

Often the target population of the survey consists of population groups with different mother tongues. To ensure the highest possible response rate and understanding of the questions, different language versions of the questionnaire are often prepared. These language versions are most difficult in the mail surveys, where the language has to be selected at the time when the questionnaire is mailed out to the survey participant.

The requirements of personnel needed for face-to-face interviews are high, especially if the interviews are done at the home of the respondents. Interviewers have to travel from house to house and often stay away from home several days at a time. This increases the required

number of interviewers. For telephone surveys, fewer interviewers are needed to cover the same sample size than in the face-to-face survey. For mail surveys, the need for personnel is even less, since there is no personal contact between staff and respondents. [19]

The face-to-face surveys are most expensive to conduct, telephone surveys being a little cheaper and mail surveys the cheapest [19, 115, 116, 120].

2.8 Physical measurements

Physical measurements are the essential part of health examination surveys. However, only a very limited number of textbooks discusses the physical measurements in health survey setting [17, 27]. Therefore the guidelines how to select measurements to be used and actual measurement protocols are rare or apply to specific surveys like the WHO MONICA Project [28], NHANES [121] or Health Survey for England [122].

2.8.1 Selection of required measurements

When selecting the measurements to be used in a survey, a few basic things should be remembered. First of all, only measurements needed to answer survey objectives should be taken to minimize the burden to respondents.

Each measurement selected for the survey should be possible to be made reliably in the population level, should be possible to standardize and be ethically approved [17]. Already existing survey manuals [28, 121, 122] can be consulted to see what kind of measurements are included in other health examination surveys.

Conclusions: Measurements selected for the survey will affect the availability of the data to answer survey objectives, but they will also affect the accuracy and validity of the measurements. If measurements, which are not possible to measure accurately at the population level in field setting are selected they will provide data but the accuracy and therefore the usability of the data may be questionable.

2.8.2 Measurement protocol

Measurement protocol is a detailed, step-by-step instruction how the measurement, for example blood pressure, should be made. The protocol also prescribes the instruments to be used for the measurement, what is required from the respondent before the measurement, and how the actual measurement is conducted.

Each survey should have its own written measurement protocol for each measurement taken in the survey. These protocols can be adopted from already existing ones, which will improve the comparability of the results between studies or they can be created specifically for the survey in question. When already existing protocols are adopted, they should be carefully evaluated to make sure that they also work in this population.

The European Health Risk Monitoring (EHRM) Project has reviewed the measurement protocols of national health surveys in Europe for the measurement of major chronic disease risk factors [123]. From this review it can be seen, that different instruments and measurement protocols are used in different surveys. Since the different measurement procedures can have an effect on the measurement results, it is important to keep the same measurement protocol from survey to survey, if one wants to estimate trends or compare results with other countries.

Conclusions: The measurement protocol increases accuracy of the measurements and the comparability of the results between surveys within population and between populations.

2.8.3 Order of measurements

The order of the measurements should be carefully planned. First of all, the respondent should have as little stress and inconvenience from the measurements as possible and secondly, the accuracy of the measurements should not be compromised by the measurement order. The inconvenience to respondents can be reduced for example, if the measurements, which require taking off clothes, are done at once.

Measurements, which are most important for answering the survey objectives should be placed early in the measurement list [17]. This assures that the data are available for the most of the respondents even when some respondent will not complete the whole list of measurement, e.g. decide to leave at the middle of examinations.

Stressful procedures should be placed so that they follow the measurements, which may be affected by stress [17]. For example, blood pressure is known to be sensitive to stress. If venous blood samples are drawn before the blood pressure measurements, they may affect the blood pressure results.

Also, if the survey is a continuation of a set of surveys conducted earlier, the measurement order should remain same [17]. This will increase the comparability of the results.

Conclusions: The order of measurements may affect the accuracy and the comparability of the results as well as item non-response.

2.8.4 Timing of the measurements

When planning the timing of the measurements, there are two issues to take into account, daily and seasonal variation of the measurements, and the requirements for the individual measurements.

Many physical measurements have seasonal variation. When a survey lasts throughout the year, the seasonal variation is not usually a problem. The problem occurs if some population groups are examined during the early part and some other population groups during later part of the survey. Then the results between two population groups are not comparable. [17]

When the survey lasts only a few months, the seasonal variation is not a problem. However, repeated surveys should always be conducted during the same months of the year [17].

Some measurements have special requirements, like fasting before sample collection. For the respondent it may be easiest to fast over night and come to the measurement the first thing in the morning [17]. If only a few hours of fasting is required, then respondents can be asked to fast also during the day and come to the survey measurement in the afternoon or evening.

Conclusions: The timing of the survey and the timing of individual measurements may affect the accuracy and the comparability of the results.

2.9 Manual of operations

Each survey should have a manual of operations. The manual of operations is a detailed documentation of the entire survey process. The manual of operations should include [35]:

- Background of the survey.
- Objectives and aims of the survey as well as the hypothesis to be tested.
- Description of the sampling frame and sample selection.
- A copy of data collection forms (questionnaires and forms in which information from the physical measurements are collected) with detailed coding instructions.
- Detailed description of included physical measurements, instruments used to conduct them, and actual measurement procedures.
- Description of the requirements and training of the staff.
- Requirements for the survey site.

- Sequence and schedule for data collection.
- Quality control procedures.
- Documentation procedures.
- Data management procedures.
- FAQ; questions commonly asked by survey participants and staff members.

The background, objectives, and aims briefly describe why the survey is conducted and the research questions to which the survey is expected to provide the answer. If the survey is set out to test some hypotheses, these also should be described in the manual of operations. [35]

The definition of target population and the eligibility for the survey should be described in the manual of operations. Also, the used sampling frame, and the calculations to which the sample size is based, and the actual sampling methods should be listed in the manual of operations.

A copy of each questionnaire used in the survey for data collection as well as forms used to code results of physical measurements should be included to the manual of operations. Also, a copy of the invitation letter and the informed consent should be included to the manual of operations. In the manual of operations, detailed instructions for interviewers should be provided for each question. These instructions include coding instructions and the instructions on how to probe the respondent if needed. For the forms used to code results from the physical measurements, the coding instructions for each point should be included. [35]

Example. The coding instruction for question [28]

Are you still having monthly periods?

1 = yes, as usual
2 = yes, but irregularly
3 = no
8 = not relevant
9 = insufficient data

could be the following:

Code 1 if she has her monthly periods (as her usual pattern)
Code 2 if she has had periods within the last 6 months but not as regularly as they used to be (excluding pregnancy)
Code 3 if she has had no periods for a period of six months or more
Code 8 for men and known pregnant women
Code 9 if insufficient information is available to use the other codes.

For each physical measurement, the required instruments should be listed with detailed, step-by-step instructions how the actual measurement is conducted. For the instruments, also the calibration and maintenance instructions are needed. Examples about these can be found in the European Health Risk Monitoring (EHRM) Project recommendations [17].

Detailed requirements for staff members and their training should be included in the manual of operations [35]. The calculations for the required number of different staff members: receptionist, interviewers, measurers, laboratory technicians, etc. are needed. These calculations should include the extra staff members for possible sick leaves and other unexpected absences. Also, the manual of operations should have a detailed job description for each group of staff members and list the skills required for their jobs. Detailed training programs for each group of staff members should also be described.

The requirements for the survey site need to be listed in the manual of operations. This includes the required number of examination rooms and their size, how reception and waiting areas will be organized, where the interviews are conducted, where the staff rooms are, how and where supplies and collected samples are stored, and how the storage of participants' overcoats and other possessions are organized during the examinations. If some measurements have special requirements for the examination room, these should also be listed in the manual of operations. [35]

Example. When a balanced-beam scale is used, the floor should be level and have hard surface (not covered with carpet or any other soft material).

The sequence and the schedule of the data collection should include a detailed flow of the measurements [35] and also instructions what to do if someone refuses to take part in some measurements.

Example. A possible sequence of the measurements in the health survey is shown in *Figure 16*. It is assumed that the invitation letter and the questionnaire are mailed and the participant is asked to complete the questionnaire at home before coming to the survey site.

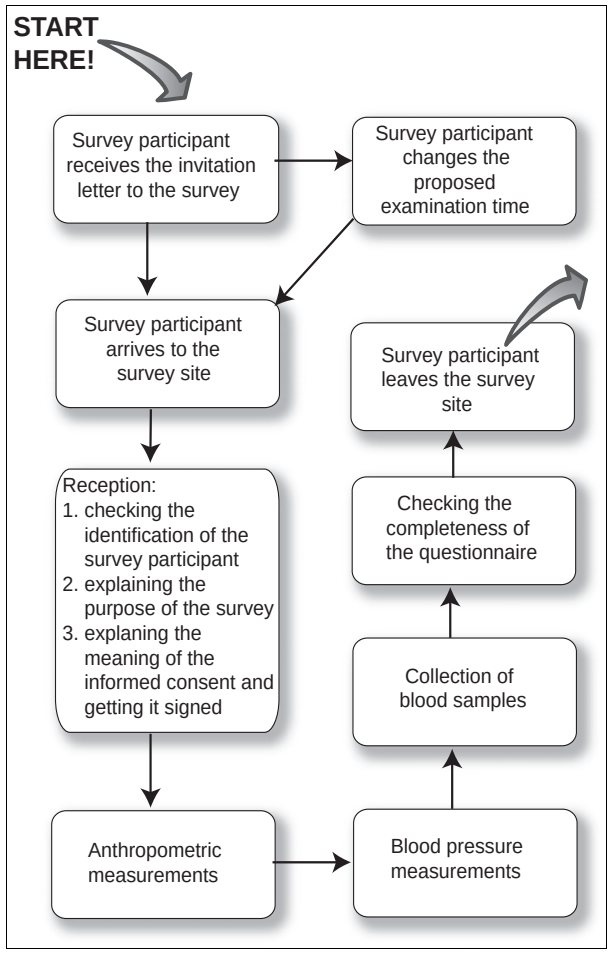

Figure 16. An example about the survey sequence

Quality control measures should be included to the manual of operations [35]. There should be a description of the quality control measures conducted during the survey and how results are passed to the survey staff. Also, the implications of the quality control results should be described in the manual of operations, i.e. when staff members are required to go through re-training/re-certification if quality control results show some problems.

It is also useful to list some basic rules of documentation of survey process in the manual of operations as well as the basics of the data management issues.

The manual of operations will be the first tool each survey staff member should consult if they have questions or problems during the survey. It should also contain a section that provides answers to the most frequently asked questions (FAQ) by survey participants and staff members [35]. This part of the manual of operations can be constructed based on the experience from the pilot study and questions raised during the training. In an ideal situation, this part of the manual of operations is an interactive part, which keeps growing and developing during the survey when new questions arise.

The well-prepared manual of operations is the most important survey tool and the reference for the survey staff. It also serves as basis for the documentation of the survey process. Each survey has its unique features; therefore, each survey should have its own manual of operations. Examples of manuals of operations can be found for the WHO MONICA Project [124] and NHANES [121].

Conclusions: Manual of operations increases the total survey quality by documenting the agreed procedures and by providing the reference for survey staff.

2.10 Training

Training the survey personnel is the corner stone of a successful survey. Training has two main components: knowledge and skills. Knowledge means that the survey personnel know

what the purpose of the survey is, and what the data collection methods are. Skills mean that the survey personnel have required skills to follow the survey protocol and methods correctly.

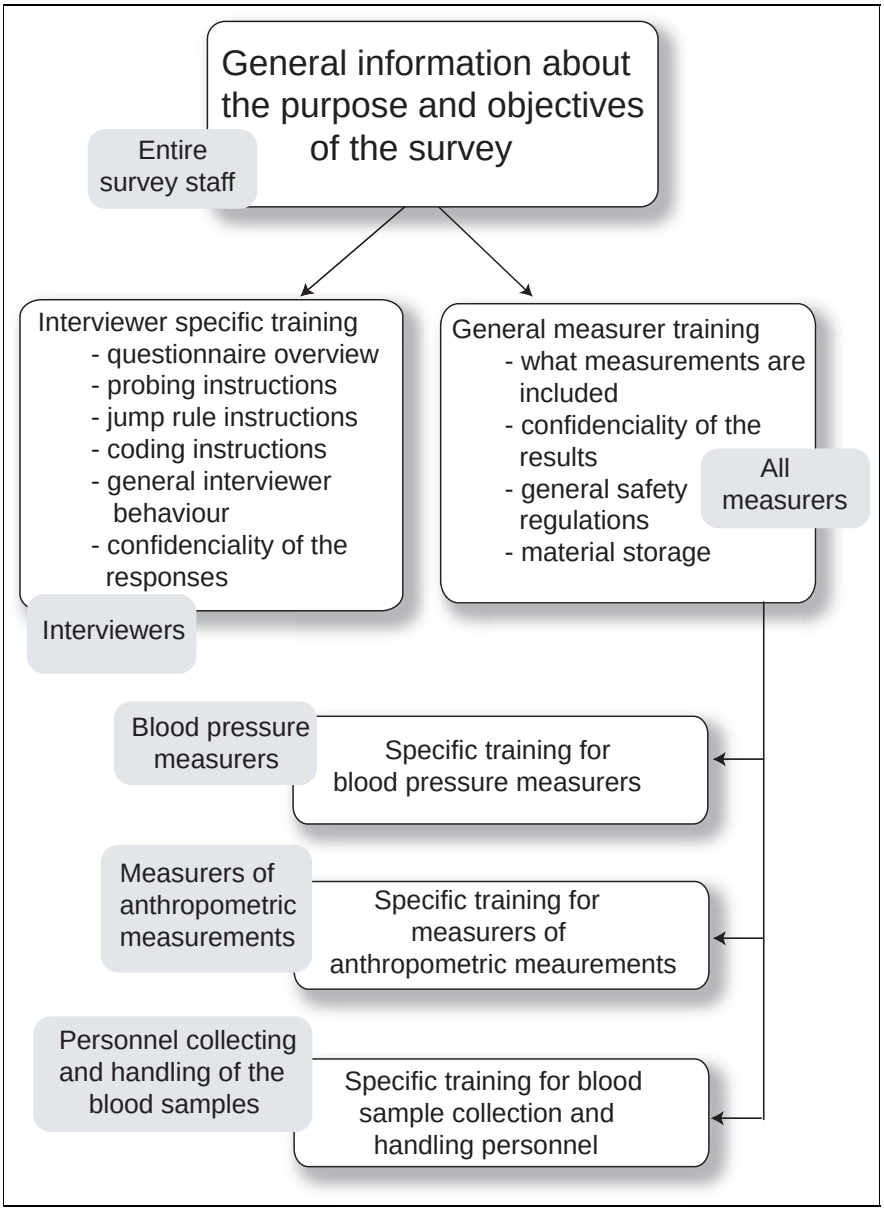

Figure 17. Training of the survey personnel

How the actual training of the survey personnel is organized depends on the final survey protocol. The manual of operations has the key position in the training, since it describes the entire survey process, includes all the forms used in the survey, and gives detailed instructions for each individual measurement [35].

Training may have a part that is common for the entire survey personnel, and covers the basic background information about the surveys. After that the survey personnel is split into smaller groups for more specific training. (*Figure 17*)

2.10.1 Training of the interviewers

The training of the interviewers includes the information about the survey, its objectives, sponsors and the role of the interviewers for data collection, and the actual training of the interviewing techniques, data coding and confidentiality [18, 25, 74]. It is important that the interviewers have some basic knowledge about the survey and its purposes since some respondents may request this information in more detail than what the introduction will provide. Emphasizing the importance of the interviewers' role in the data collection may help to increase the motivation of the interviewers.

There are several different interview technique [125]:

- no clarifications (standardized),
- scripted clarification only when asked,
- scripted clarification also unsolicited,
- paraphrase clarification only when asked, and
- paraphrase clarification also unsolicited.

Regarding the interview techniques, there are two main schools of thought. One advocates a strictly standardized interview, where the questions are read as they are and the interpretation is up to the respondent. The other stresses that the key is the meaning of the questions, in other words, the interviewer's task is to make sure that each respondent understands the

question in the same way. For simple questions, there is no difference in the results between the interview techniques but when questions get more complicated, the strictly standardized interview gives less accurate responses than others, especially when compared to the paraphrase unsolicited clarification (conversational interview). The unsolicited clarification produces more often accurate results than the clarifications only when asked. The amount of clarification that the interviewer provides has an effect on the length of the interview. The more clarification the interviewer provides, the longer the interview will take. [125]

Regardless of the interview technique, the interviewer training should include the overview of the questionnaire and what types of questions are includes. Each question should be reviewed and detailed coding and probing instructions should be provided and discussed. Also, if some questions have jump rules these should be discussed and explained. The training should include guidance in the general behaviour of the interviewer and explain the importance of anonymity and confidentiality of the respondents. [17, 18, 74]

Interviewers often tend to read the questions too fast and do not use the exact wording of the questions. During the training, the speed, clarity, and the wording of the questions should be rehearsed. [18, 19, 74]

If the interviews will be conducted using computers for data entry, i.e. the interview will be either CAPI or CATI, the training has to include the instructions related to the use of hardware and software. [126]

Conclusions: The training of the interviewers will have an effect on the accuracy and response rate of data obtained by the questionnaire. Well-trained interviewers know how to code answers and will provide adequate and correct probing and will not direct respondent to give certain answers. They also know how to persuade reluctant respondents to participate.

2.10.2 Training of the measurers

Training of the measurers is very important to ensure the validity of the measurements. The training of the measurers should include the general introduction to the purpose and objectives of the survey, and the measurement-specific training. The measurement-specific training should include the introduction of measurement instruments, how they are properly used, how they are maintained and calibrated, if calibration can be done in the field. Then the actual measurement protocol should be reviewed and explained in detail. The important part of the training is the practicing of the measurements. Each measurer should have a pre-defined number of correct measurements before they are accepted to do the actual survey measurements (certification). [17, 35]

For some measurements training protocols exists in the EHRM recommendations [17] and in the NHANES protocols [121].

Conclusions: The training of measurers will affect the validity of the measurements.

2.11 Pilot study

A pilot study is usually conducted to evaluate the entire survey process and to obtain the additional information needed for the planning of the actual survey [5, 20, 35]. For the pilot study, a sample of 100 to 200 subjects is used [20].

During the pilot study, the questionnaire is evaluated in regard to the length of recall period, clarity of concepts and definitions, the question wording and the response alternatives, the sensitivity of topics, the questionnaire layout, the choice of administration mode, and the respondents' burden. Additionally, for each measurement, the feasibility and the measurement protocols are evaluated, as well as the wording of confidentiality pledge, data entry and management procedures. [20]

Conclusions: The pilot study will help to detect and correct possible shortcomings of the survey process before the actual survey takes place. Even though pilot studies result is costs, they will, in long run, save money, as they help to improve the quality of the actual survey.

2.12 Response rates

Response rate is one indicator of the survey quality. It represents the proportion among the people initially selected to the survey who then answer the questionnaire and take part in the physical examinations. High response rate indicates good coverage of the target population whereas low response rate indicates problems with the coverage.

When a person who was selected to the survey does not answer the questionnaire or does not take part in any of the physical examinations, he/she contributes to the unit non-response. Persons who answer at least some of the questions and takes part in some of the physical examinations contributes to the item non-response. [2]

Problems with locating, solicitations, or data collection may cause non-response. Problems with locating mean that the survey organizers do not have accurate information about the person's address or otherwise fail to make contact with the person. This is usually due to outdated contact information in the sampling frame or temporary absence from the address. The problems with solicitation mean that the survey organizers can locate the person but the person refuses to participate. Reasons for non-participation can vary from not having time to not being interested or not being willing to provide personal information. The problems with data collection are related to the questionnaire and the physical examinations. A person may take part in the survey but he/she does not provide full information for all the survey questions and the physical measurements. Persons may feel that the questions are too personal or too difficult to answer or that some physical measurements are uncomfortable or threatening and, therefore, they will not take part in them. [2]

Nowadays there is a general decrease in survey response rates. People become more reluctant to participate or are less at home, i.e. they are more difficult to contact. There are differences

in the response rates between countries indicating some cultural differences in the attitudes and behaviours towards surveys. [127]

People tend to respond to surveys always in the same manner, they either are always respondents, or they are always non-respondents or refusers [128]. In general, non-respondents tend to be younger than respondents and they are more likely to be single, have lower education and have a worse health profile [129, 130]. People with physical problems are more likely to participate in health surveys [131].

Table 9. Features of intermediate and late respondents, initial and passive refusers, and people who are hard to contact in comparison to early respondents

Response status	Age group	Sex	Marital status	Educational level	Health status
Intermediate respondents	Young			Low	
Late respondents	Young			Low	
Initial refusers	Old	Men		Low	
Passive refusers	Young	Men	Single		
Hard to contact	Young	Men	Single		Excellent

When comparing the early respondents, intermediate and late respondents, initial and passive refusers, and those who are hard to contact, one finds some distinct differences between these groups (*Table 9*). The intermediate and late respondents tend to be younger, and have lower education than the early respondents [132]. Also, the initial refusers have lower education than the early respondents but the initial refusers tend to be older than the early respondents [108, 132]. The initial refusers and the early respondent have similar health history [132]. The passive refusers are more likely to be young men who have never been married and have high socio-economic status. Persons who are hard to contacts are more likely to be young men who have never been married and who have excellent perceived health. [108]

Overall non-response rate is a measure of the survey data quality, reflecting the possible difficulties of the data collection process and the skills of survey organizers [133]. The response rate per individual interviewer can be used as the quality indicator for that interviewer's work. [2]

Conclusions: The response rate reflects the coverage of the sample. Low response rate indicates that the coverage is low, i.e. the results may not be representative for the target population.

2.12.1 Calculating response rates

Simple response rate is usually calculated as the proportion of the survey sample that participates in the survey. Sometimes it may also be of interest to calculate some other fractions of the survey sample.

An often used classification of the initial survey sample is described in the *Figure 18*. In this classification, the originally selected survey sample can be divided into eligibles and not eligibles. The eligibles are those, who fulfil the survey definition of eligibility (see *Chapter 2.4.1*); the remaining is usually classified as not eligibles. Among eligibles, persons can be either respondents or non-respondents. Usually, the respondents are those who complete the questionnaire and/or take part to the physical measurements. The non-respondents are those from whom one cannot obtain any information: questionnaire or physical measurements. The respondents can still be divided into two groups: depending on whether they provide complete or partial responses. The respondents with complete responses have provided data for the questionnaire and have taken part in the physical examinations, while the respondents with partial responses may have filled in the questionnaire but not taken part in all the physical examinations. Since these respondents with partial responses have provided data for a major proportion of the survey measurements, they are usually not considered non-respondents. The non-respondents are usually those who either refuse to participate or just for some reason cannot be contacted to verify their status.

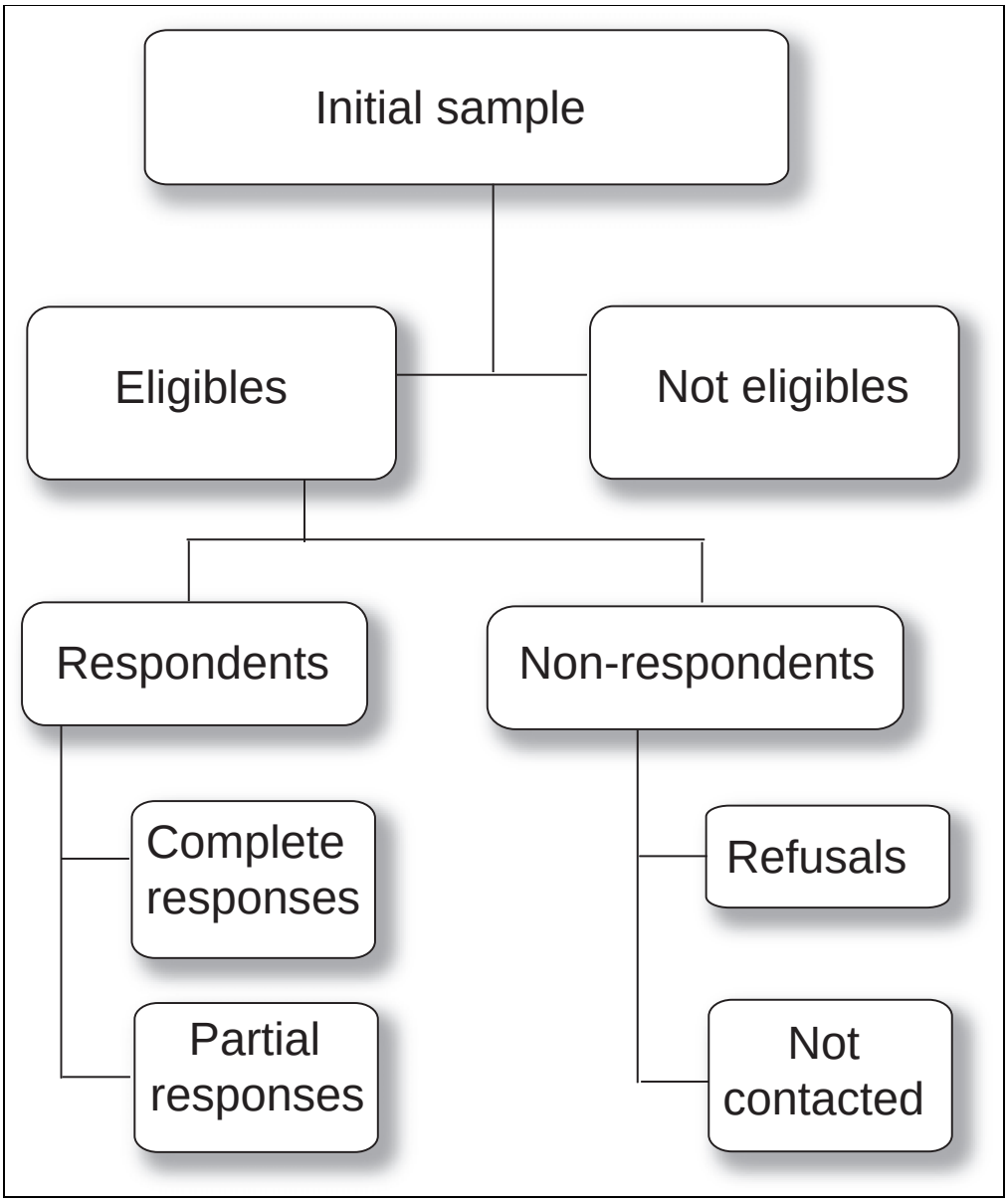

Figure 18. Classification of the initial survey sample

From the classification of the survey sample given in *Figure 18*, the response rates, the co-operation rate, the refusal rate, the non-response rate and the non-contact rates can be calculates [5].

Response rate: $\frac{respondents}{eligibles}$ or alternatively $\frac{respondents}{eligible - not\ possible\ to\ contact}$.

Co-operation rate: $\frac{respondents}{respondentes + refusals}$.

Refusal rate: $\frac{refusals}{eligibles}$.

Non-response rate: $\frac{non - respondents}{eligibles}$.

Non-contact rate: $\frac{not\ contacted}{eligibles}$.

Response rate is a measure of the success of the locating and solicitation steps, while the refusal rate measures how well survey organizers were able to get located persons to participate in the survey, i.e. the success of solicitation. [2]

2.12.2 What is high enough response rate?

The response rate should be high enough to provide the results, which can be generalized to the entire target population without bias. When the non-respondents are similar to the respondents in all aspects (age distribution, sex, socio-economic status, etc.), event a low response rate will not bias the results but will only decrease their precision [134].

Several studies have found marked differences between respondent and non-respondents. Non-respondents tend to be more often younger, single, men, and have lower education than are respondents. [25, 135-142] These observed differences between respondents and non-respondents will cause bias in the results if the response rate is not high enough.

Previous studies have suggested that in mail surveys, the response rate should be at least 50%-60% in the first round and at least 60%-75% after reminders [2, 25, 74]. In interview surveys, the response rate should be around 70% [2, 25]. These recommendations are based on the assumption that the respondents and non-respondents do not differ from each other. If there is difference between the respondents and non-respondents, the response rate should be 90% or higher [143]. Since the studies about the respondents and non-respondents seem to systematically find differences between these two groups, the general recommendation for the response rate should be closer to 90% than to 70%.

Conclusions: High response rate will improve the accuracy and the precision of the estimates.

2.12.3 Increasing response rates

The first contact with people selected to the survey often determines their decision to participate or not. Therefore, it is import to consider how the first contact is made. Some small adjustments can possibly increase the response rate. Several studies have shown, that a pre-notification letter, telling people that they will soon be contacted about participation in the survey, will increase the response rate [2, 19, 20, 142, 144-146] by 9-47% [145, 146]. In case of personal interviews, an advanced phone call to schedule the personal interview holds potential to reduce the survey cost and poses only a small risk of increasing the refusal rate [147].

Reminders are known to increase the response rate [2, 19, 20, 144, 145]. Reminders may increase the response rate by 11-35% [145].

In telephone surveys, increasing the attempts to place a call to the numbers that do not answer may increase the response rate [2, 148]. Also, leaving a message to the answering machine has been shown to increase the response rate by 15%. A message on the answering machine seems to work especially well with younger age groups and women in the upper socio-economic class. [149]

In mail surveys, one possible method in trying to increase the response rate is the telephone call to non-respondents requesting them to participate [25]. On the other hand, mentioning the possible telephone contact in case of non-response already in the invitation letter may even decrease the response rate [142].

In mail surveys, the simple modifications of the questionnaire mailing may increase the response rates. It has been shown that the use of the first class postage with real stamps increases the response rate [2, 19, 20, 145, 150]. First class postage may increase the response rate by 9-10% [145, 150] and the use of real stamps by 5-32% [145]. Also the colour of the questionnaire paper, the layout of the questionnaire, the official sponsorship, the personalized correspondence, the title under the sender's name, and the stamped return envelope are known to have an increasing effect on the response rate [19, 20].

A respondent-friendly questionnaire is short, clear and easy to comprehend and will increase the response rate [19, 20, 142].

Another way to increase response rate is to offer different kinds of incentives [2, 19, 25, 146, 150]. These incentives can be simple pens or a lottery for a weekend in a spa, short cruises, or money, etc. People from lower income groups are found to be more receptive to monetary incentives than people from high income groups [150]. The response rate also seems to correlate positively with the magnitude of monetary incentive [146].

There are no studies about the ways to increase the response rate in health examination surveys. Of course, those basic principles about the first contact and the use of incentives will still apply to health examination surveys because they usually include also a questionnaire. But how to get people interested in the physical examinations of the health surveys is a problem. Physical examinations will take time and people often have to take time off the work to participate to the examinations. Some measurements may be slightly uncomfortable for some people. These are often reasons why people do not participate in health examination surveys. If the only value gained from the survey is information about their blood pressure and cholesterol measurements it is not enough to get people interested in participation.

2.13 Quality control during the survey

Quality control during the survey is used to guarantee the proper use of survey instruments and measurement techniques throughout the survey, i.e. minimize the measurement error due to faulty use of the instruments. When possible problems are noticed already during the survey, they can be corrected before they would cause irreversible errors to the survey data.

2.13.1 Interviewer monitoring

Quality control of the interviewer work can be done by monitoring the interviews, i.e. a monitor listens to the actual life interview or on a tape, or by re-interviewing the respondents [19, 35, 74]. When the interviews are monitored, either life or from the tape, the respondent has to be informed about that [74].

Re-interviewing can be used to check that the interviews actually took place and that all the questions were asked. This is done to minimize the possibility that the interviewer has fabricated the interviews. [19, 74] Also, just counting the number of interviews each interviewer has conducted during the day may reveal possible problems.

Conclusions: Monitoring of the work of the interviewers helps to ensure that they really conduct the interviews and all the questions are asked and asked using correct wording and probes.

2.13.2 Measurement quality control

Quality control measures for the measurements, during the survey depend on the specific measurement. One way to ensure the accuracy of the measurements is that two measurers measure the same individual and these results are compared. In addition, during site visits by survey supervisors, the measurement instruments and the techniques can be evaluated [35]. There are also a number of ways to monitor the measurement results and evaluate their accuracy. The European Health Risk Monitoring (EHRM) project has proposed a quality control procedures for blood pressure measurement, the blood collection, and the anthropometric measurements [17].

Conclusions: Monitoring of the measurement quality during the survey helps to ensure the proper use of measurement instruments, which increases the accuracy of the measurements.

2.13.3 Feedback to survey personnel

Survey personnel should receive feedback about the quality control results as soon as possible, preferably within a few days. This helps to correct possible errors quickly and improve the performance. In case where just informing the interviewer or measurer about their problems does not produce improvement, re-training may be required [35].

The feedback about the quality control results should be constructive but in some cases may cause even more problems. It should be carefully thought how feedback is given.

> *Example.* A survey nurse measuring blood pressure tends to have strong zero preference, which is picked up during the quality control procedures when checking the terminal digit preferences of the blood pressure measurements. The nurse is informed about this tendency and is asked to correct her behaviour and try to measure the blood pressure according to the instructions given during the training. The nurse tries to correct her measurements by avoiding the zero

preference. The next quality control round reveals that now she has a preference for another even digit and she avoids the zero as terminal digit. In this case, the actual problem has not been fixed by the feedback of the quality control results but has just changed its form.

Nowadays, when the Internet is getting more common and is more easily accessible in many countries, the survey organizers should consider the use of the Internet for distribution of the quality control feedback and the additional instructions for the survey field personnel.

Conclusions: Fast feedback about the results of survey quality control helps to correct possible errors quickly and minimizes the error in collected data.

2.13.4 Quality control of data management

During the surveys, a lot of information about the respondents is collected in the field by questionnaires and physical measurements. As part of the quality control system, there should be a mechanism that verifies that all the information collected from the respondent has been transferred to the main database, and that the transferred information is exactly the same as the information collected in the field.

Conclusions: Quality control of the data management system makes sure that no information about the respondent collected in the field gets lost during the data transfer from one place to another.

2.14 Retrospective quality assessment

Retrospective quality assessment reports are prepared after the survey to document how the different data items were collected, what protocols were used, and whether the methods used in the field differed from the ones prescribed by the protocols. Retrospective quality

assessment reports are also used to assess the availability and coding of data item by item. This kind of examination of achieved level of standardization helps to evaluate the representativeness and the comparability of results, especially in the case of multinational surveys. The retrospective quality assessments also serve as documentation for those using the data and reporting the results based on them. From the retrospective quality assessments, the organizers of future surveys can learn a lot about performance of the measurement instruments, measurement protocols, and general organization of the surveys.

The WHO MONICA Project [6, 7] prepared a total of 13 quality assessment reports from its survey data [8-10, 13-16, 151-156]. These MONICA quality assessment reports are the only existing published retrospective quality assessments and therefore, this Chapter on the retrospective quality assessment is solely based on the MONICA experiences and results.

For health examination surveys like MONICA there are two main types of quality assessment reports: 1) the quality assessments for measurement like blood pressure, cholesterol, or anthropometric measurements [10, 151-154] and 2) the quality assessments for the questionnaire items [13-16, 155, 156]. Additionally, in MONICA there is one report specifically for the representativeness of the sample through sampling frames and response rates [9] and another one for the age of survey participants and the timing of the survey [8].

The components included to the different kinds of quality assessment reports are briefly listed here. The reports used here are examples from the multinational WHO MONICA Project where the main objective was to be able to compare the results in time within populations but also to draw conclusions between populations. Each of the four types of quality assessment report used in MONICA has specific components but also each of them has one common component, the quality score. This, artificially defined score summarizes the data quality for each survey item in quantitative format, which can be used to compare the data quality between and within the populations.

Conclusions: Evaluation of the actually used survey methods in comparison with the methods listed in the manual of operations provides the information about the quality, reliability, and the comparability of the results.

2.14.1 Quality assessment report for measurements

The content of the quality assessment reports for the physical measurements varies since each measurement has its special features but the purpose remain the same, to document the used measurement instruments and procedures and the achieved data quality.

The blood pressure quality assessment report [10] documents the blood pressure measurement devices that had been used, the measurement procedures that had been employed, how measurers were trained and certified, and what quality control measures were taken during the survey. Using the collected blood pressure data; the difference between two measurements taken, the terminal digit preference, the proportion of identical readings, and within survey time trends were examined.

For blood pressure measurements, information about the arm circumference, the room temperature during the measurement, the time of the day of the measurement and the measurer are recorded. Availability of this information is documented and the mean and standard deviation for the time and temperature are calculated. [10]

The summary score for blood pressure measurement is a combination of scores for the proportion of incomplete measurements, the proportion of odd readings, the terminal digit preference, the proportion of identical results in duplicate blood pressure measurements, and the time trend within survey. A separate score for the blood pressure trends between surveys was derived by looking at the over all summary score, the changes in device, cuff, posture, order of measurements and survey months between the surveys. [10]

Quality assessment reports for total cholesterol as well as for HDL cholesterol have very similar structure [152, 153]. Just like for blood pressure quality assessment, the quality assessments for total cholesterol and HDL cholesterol document the used devices, the measurement protocols, the training and certification of the measurers, the handling of the samples, and the results of external quality control. The summary score for total and HDL cholesterol was derived from the variance score, the bias score, and the coverage score, all relating to the external quality control [152, 153].

For anthropometric measurements there were separate quality assessment reports for weight and height measurements [151] and the hip and waist circumference measurements [154]. Both quality assessments documented the used devices, the measurement protocols, and the accuracy of the measurement. From the collected data, the terminal digit preference was checked for each measurement. The summary scores for anthropometric measurements were based on the proportion of terminal zeros, the proportion of zeros in the second last digit and to the proportion of missing data.

2.14.2 Quality assessment report for questionnaire items

For all the questionnaire items, the structure of the quality assessment reports was very similar [13-16, 155, 156] and same information is repeated in the several quality assessment reports. In each report, the way in which the questionnaire was administered is documented (self-administration or interview). Reports also include information about the comparability of local questions with MONICA standard questions, the proportion of missing data for each data item, the local coding practices for each data item, and finally the summary quality score and in most cases also the trend score for all three surveys.

The summary score [13-16, 155, 156] is usually based on the proportion of insufficient data (i.e. missing information among respondents), the comparability of locally used questions with MONICA standard questions, and in case question included the answer alternative “Uncertain”, the proportion of uncertain answers.

2.14.3 Quality assessment of sampling frames and response rates

The purpose of the quality assessment of sampling frames and response rates [9] is to document information about the used sampling frames, the definition of eligible and non-respondents, the obtained response rates (overall and item response rates), the availability of non-respondent data, and the reasons for non-response.

For the sampling frame, the information about the type of sampling frame (population register, general practitioners list, electoral roll, etc.), the age of sampling frame, the proportion of those not eligible to the survey, and the proportion of not possible to contract is recorded. For the eligibility and non-respondents the used definition are documented.

MONICA used two different definitions for the response rates. The response rate A is defined as a number of respondents divided by the eligible sample size, and the response rate B is defined as a number of respondents divided by eligible sample size minus the respondents who were not possible to contact. Item response rate is calculated for the systolic blood pressure, total cholesterol, BMI, and smoking.

In MONICA, it was recommended to collect as much information from non-respondents as possible and basic data items were also recorded for the non-respondents. The availability of the non-respondent data is documented as well as the reason for non-response (not possible to contact, temporarily out, medical reason, not interested, or other refusal).

Sampling fractions, the proportion of population in the sample is calculated within each population and survey for the reporting units and 10-year age groups. Different quality scores are calculated to assess the quality of the sampling frames, and the agreement between the individual level and aggregated data.

2.14.4 Quality assessment of age of survey participants and survey time

The purpose of the quality assessment of the age of survey participants and the survey time [8] is to document the information about the age stratification of the sample, the data of date of birth and date of examination, the sex and age distribution of the survey participants during the survey months, the age distribution within 10-year age groups, and the information about the survey period and possible seasonal variation between the surveys.

In the WHO MONICA Project, it was recommended to draw samples using the age stratification by 10-year age groups. Therefore, information about the sampling age strata was collected.

The completeness of the date information is checked for date of birth and date of examination. This is done by counting in each survey/population the number of respondents with complete date information (date, month and year), the number with missing dates, and the number where only the year is known. For the date of examination, the distribution of survey respondent by sex and age group in each survey month is checked to see whether participation was evenly distributed to avoid seasonal variation between the sex-age groups.

When the sample was stratified by age, the information about the age strata from which the person was sampled is recorded. Based on the data from the date of birth and date of examination, the respondent's exact age in full years at the time of examination is calculated. Comparing the distribution of the sampling age groups and the actual age groups at the time of the examination gives additional information about sampling. The age used for sampling can be defined as the age at the sample selection, the age at the beginning of the survey period, the age at the middle of the survey period, or the age at the end of the survey period. If the sample is drawn long before the actual survey takes place and the age is defined as the age at the time of sample selection, the population participating in the sample does not correspond to the target population during the survey period. The distribution of age within the age groups is examined by checking how much the mean age in each 10-year age group deviated from the expected mean.

The survey period for each survey is documented at the accuracy of month and year, and the possible seasonal difference between the surveys within the populations is investigated.

2.15Reporting survey results

Reporting of the survey results is the aim of each survey. The survey should be able to answer the survey objectives. The reporting of the survey results can be based on the following three questions:

1. To whom to report?
2. What to report?
3. How to report?

Each of these questions is equally important and they are interrelated, therefore they should not be considered in isolation.

The health survey results can provide information, which is of interest to many different groups; researchers, health professionals, politicians, health educators, different health related organizations, and also the general public. Each of these groups has different needs for the health information and different abilities to process the available information. [157]

The reporting of the results of the health survey should be adapted to the needs of the target group. This means that the level of details, format of presentation, and amount of additional explanations depends on the target group for whom the results are meant. Researchers and health professionals usually can draw their own conclusions from the presented numbers but for the general populations, the meaning of the numbers should be explained. Also, the big, extensive tables may provide a lot of information and are therefore interesting for the researchers but for politicians and the general public, simple, informative figures are easier to understand. [157]

The question *"What to report?"* appears also in the Declaration of Helsinki. The Declaration of Helsinki states that the researcher is responsible for ensuring that the published results are correct and also that negative findings are published or otherwise made publicly available [34]. The correctness of the results means, that the data are of high quality, the results are representative for the target population, and the data are analysed using the proper methods. To fulfil these criteria, the researcher should report information about the target population, how representative the sample was for that target population, and how representative the obtained data are for the sample. Additionally the information about the measurement instruments and possible measurement errors should be reported. [25]

The reporting of the survey results is the highlight of the survey. When reporting the results, the researcher should be critical about the results and also discuss the possible biases in the data. Admitting that there may be some bias in the results due to low response rate or observed systematic measurement error does not mean that the results are wrong and unusable. On the contrary, it means that the researcher has evaluated the quality of the survey data and is aware of the potential bias and therefore can take that into account when interpreting the results.

Conclusions: The reporting of the health survey results should be adjusted for the target group and include information about the representativeness of the results as well as about the accuracy of the results. Reporting of the data quality information will increase the reliability of the results.

2.16Documentation

In the existing survey literature, there is much discussion about methods how to conduct the survey but no mention is made about the importance of the documentation. The proper documentation of the entire survey process may seem self-evident but in many cases it is forgotten.

The documentation of the survey process should cover at least:

- the objectives and rational of the survey,
- the definition of the target population and the eligibility within the target population,
- the description of the used sampling frame and how the actual sample was drawn,
- the determination of the sample size,
- the original references to the survey instruments which are adopted from other surveys or national/international recommendations,
- the description and validation of new survey instruments developed specifically for this survey,
- the requirements and selection of the survey personnel,
- the training program and results of evaluation of the personnel,
- the actual implementation process of the survey,
- the quality control processes carried out during the survey and their results,
- the methods used for data collection, transfers, coding and editing,
- the data analysis, and
- the reporting of the results.

The documentation of the target population, sampling frame, and sampling methods will help to assess the representativeness of the survey results to the target population. The documentation of personnel training, the survey instruments and processes in the field together with the quality control information collected during the survey can be used in the evaluation of the data quality. The documentation of the data management issues is essential for the verification of the correctness of the data. Each change made to the original data should be documented so that one is able to explain the possible differences between the original data and the final, edited data. The documentation of the data analysis and the reporting of the results are linked, since there should be clear documentation about which analysis is used for which reports. It may be that at some stage, someone asks how the reported results were calculated and then the proper documentation is of essential help.

The documentation should be organized so that it is easily accessible for anyone who may need it during data analysis and reporting or for some other purposes later on. In what format the documentation exists does not make any difference, as long as it does exist. One alternative format for the documentation is WWW, either intranet or the Internet, which allows easy distribution of the documentation to all who need it.

Conclusions: Proper documentation helps to identify representativeness of results to the target population, used procedures, data quality, and the accuracy of the reported results.

2.17Data management

Another topic, generally omitted in the survey literature is data management. The EHRM has listed key points of data management, which should be taken into account when conducting a survey [17].

Well-organized data management ensures that the available data are complete, correct, and verifiable, and that the data analyses are done using correct data, without errors, and that the confidentiality of the data is secured. The completeness of the data means that the data collected from the respondents are not lost during the process. The correctness and verifiability means that the data originally obtained from the respondents are the ones in the database unless there is some documented justification why they have been changed. [17]

Data management starts already at the planning stage of the survey. When a sample is selected, each subject should be assigned a unique identification code, which follows the subject throughout the survey. This also means that those found ineligible and to be non-respondents are in the database with proper identification about their response status. [17]

Data management also includes error checking of coded data. This error checking should include at least the checking of close-ended answer alternatives i.e. that only those alternatives listed in the questions are used. Also, if the questionnaire has jump rules or sets of

questions meant only for some respondents, crosschecking between the questions should be conducted. [17]

> *Example.* A questionnaire has a set of questions about smoking. The first question asked whether the person ever smoked, and if not, it instructs to pass all following smoking questions and to move to the next set of questions. In a case like this, it should be checked that all respondents who have reported that they have never smoked have not answered any of following smoking questions.

Conclusions: Well-organized data management is part of documentation and ensures the correctness of obtained results.

3 AIMS OF THIS STUDY

3.1 General aims

The general aim of this thesis is to study the process of the population health survey in relation to data quality. The aim is to identify potential sources of error in the survey process, which could reduce the quality of survey outcome and to study the possible effects of these errors on the results.

3.2 Specific aims

The general aims are approached through special topics of the survey quality.

I. The problem of non-response in health surveys is investigated at two levels. At first, the differences between the respondents and non-respondents are studied and then, the effect of non-response on the population estimates in both cross-sectional and trend analysis is studied. (Study I)

II. The quantification and documentation of the obtained data quality is discussed by demonstrating how different sources of survey errors can be quantified with quality scores, and how these scores can be used in the analysis to assess the robustness of the results in relation to data quality. (Studies II, IV)

III. The accuracy of the measurement results in relation to the measurement protocol is studied through the example of the total cholesterol measurement and how the changes in the pre-analytic measurement procedures may change the true population total cholesterol distribution. (Study III)

IV. The effect of selected indicators on the results is studied through the example of the prevalence of hypercholesterolaemia, where the sensitivity of the results to the changes in the definition of indicator is investigated. (Studies V, III)

4 MATERIAL AND METHODS

In all five studies, data from the World Health Organization (WHO) Multinational MONitoring of trends and determinants in CArdiovascular disease (MONICA) Project [7] risk factor surveys were used. The studies IV and V report the actual MONICA results and the other studies (studies I-III) use the MONICA data only as basis for the hypothetical demonstrations. In studies I and III, also simulations were used to demonstrate the quality issues.

4.1 The data from the WHO MONICA Project

4.1.1 Populations

The WHO MONICA Project was established in the early 1980's to monitor changes in and determinants of cardiovascular disease at the international level. The first main hypothesis of the project was to test whether there was any relationship between 10-year trends in classical cardiovascular risk factors and 10-year trends in coronary heart disease incidence rates. [7] Altogether, 37 populations from 21 countries on four continents participated in the project (*Figure 19*).

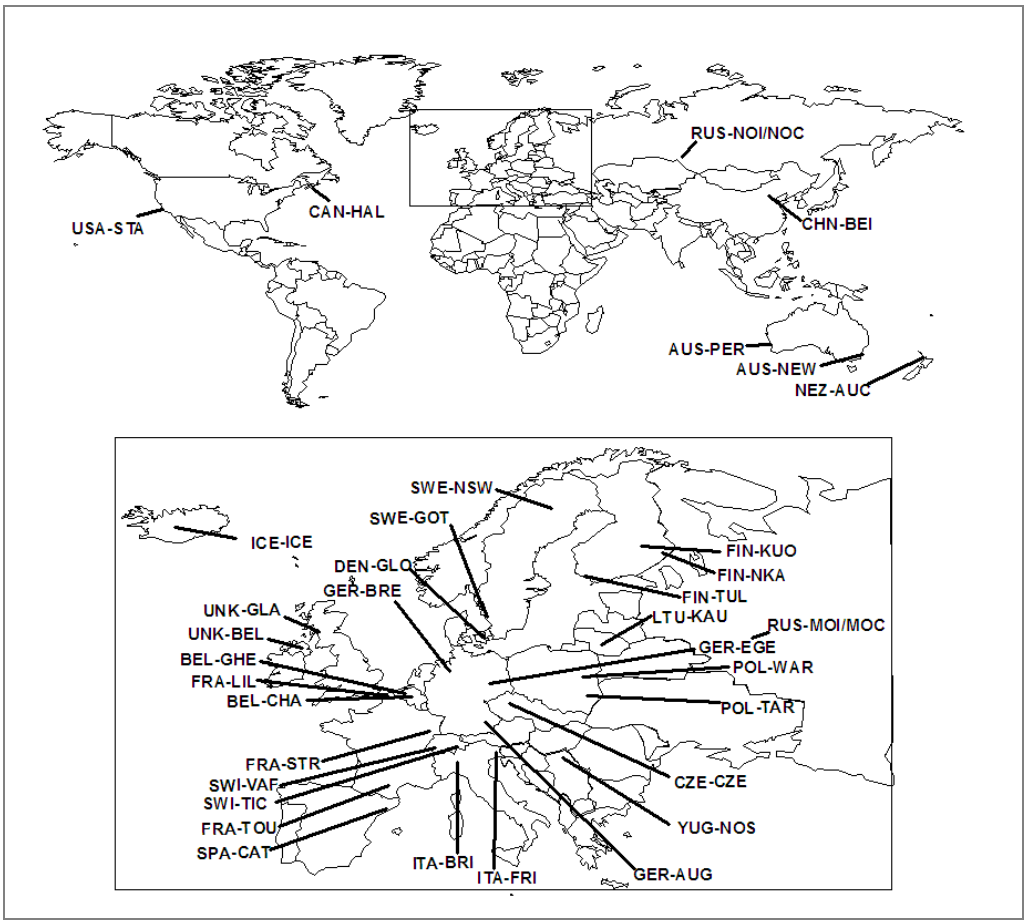

Figure 19. WHO MONICA Project populations

Risk factor surveys were conducted in each population at the beginning of the study period (initial survey) and at the end of the study period (final survey). An optional middle survey was conducted in most of the populations. The study populations covered the age group 35-64 years, but most populations were extended to cover the age group 25-64 years. The actual number of the populations used in the different studies varies due to the differences in the data availability between populations. (*Table 10*)

Table 10.WHO MONICA Project populations, and age groups covered by risk factor surveys

Country	Population	Abbreviation	Used in study	Risk factor surveys	
				Age group	Number of surveys
Australia	Newcastle	AUS-NEW	I-V	35-64	3
	Perth	AUS-PER	II-V	25-64	3
Belgium	Charleroi	BEL-CHA	II-V	25-64	3
	Ghent	BEL-GHE	II-V	25-64	3
Canada	Halifax County	CAN-HAL	I-V	25-64	2
China	Beijing	CHN-BEI	I-V	25-64	3
Czech Republic	Czech Republic	CZE-CZE	I-V	25-64	3
Denmark	Glostrup	DEN-GLO	I-V	25-64	3
Finland	Kuopio Province	FIN-KUO	I-IV	25-64	3
	North Karelia	FIN-NKA	I-IV	25-64	3
	Turku/Loimaa	FIN-TUL	I-IV	25-64	3
France	Lille	FRA-LIL	I-V	35-64	2
	Strasbourg	FRA-STR	I-V	35-64	2
	Toulouse	FRA-TOU	I-V	35-64	2/3[#]
Germany	Augsburg Rural	GER-AUR	I-IV	25-64	3
	Augsburg Urban	GER-AUU	I-IV	25-64	3
	Bremen	GER-BRE	I-V	25-64	3
	East Germany	GER-EGE	I-V	25-64	3
Iceland	Iceland	ICE-ICE	I-V	25-64	3
Italy	Area Brianza	ITA-BRI	IV	25-64	3
	Friuli	ITA-FRI	I-V	25-64	3
Lithuania	Kaunas	LTU-KAU	I-V	35-64	3
New Zealand	Auckland	NEZ-AUC	II-IV	35-64	2
Poland	Tarnobrzeg Voivodship	POL-TAR	I-V	35-64	3
	Warsaw	POL-WAR	I-V	35-64	3
Russia	Moscow Control	RUS-MOC	II-V	35-64	3
	Moscow Intervention	RUS-MOI	II-V	35-64	3
	Novosibirsk Control	RUS-NOC	II-V	25-64	3
	Novosibirsk Intervention	RUS-NOI	I-V	25-64	3
Spain	Catalonia	SPA-CAT	I-V	25-64	3
Sweden	Gothenburg	SWE-GOT	II-V	25-64	3
	Northern Sweden	SWE-NSW	I-V	25-64	3
Switzerland	Ticino	SWI-TIC	I-V	35-64	3
	Vaud/Fribourg	SWI-VAF	I-V	35-64	3
UK	Belfast	UNK-BEL	I-V	25-64	3
	Glasgow	UNK-GLA	I-V	25-64	3
USA	Stanford	USA-STA	II-V	25-64	3
Yugoslavia	Novi Sad	YUG-NOS	I-V	25-64	3

[#] Middle survey for men only, i.e. 2 surveys for women and 3 for men

4.1.2 Data collection and quality control

4.1.2.1 Data collection

The risk factor surveys were sample surveys from the target populations. In most cases, the sample was either a simple random sample or a stratified random sample. Some centres used multistage sampling. [9]

Each centre was responsible for its individual surveys. Centres were instructed to follow the survey protocol given in the MONICA Manual [28] as closely as possible. It was noted that sometimes centres had to deviate from the MONICA Manual to maintain continuity with their previous surveys or due to other local requirements. All these deviations were reported in the retrospective quality assessment reports (http://www.ktl.fi/publications/monica/).

Brief descriptions of the methods of the total cholesterol measurement and recommended questions for awareness, treatment and measurement of high cholesterol are given here. More detailed protocols for these and other survey items can be found in the MONICA Manual [28, 158].

In the WHO MONICA Project, it was recommended that serum instead of plasma should be used for the total cholesterol determination. Blood sample should be drawn from the subject in sitting position without prolonged venous occlusion. The Manual also instructed in detail about the handling and storage of the samples. [158]

Questions used in the WHO MONICA Project to collect information about the awareness and treatment of hypercholesterolaemia and the measurement of total cholesterol were [28]:

> *HICH*: **Have you ever been told by a doctor or other health worker that you have high blood cholesterol?**
> 1 = yes
> 2 = no (if no, record 8 in items *CHDT* and *CHRX*)

9 = insufficient data

CHDT: **Are you on a special diet prescribed by a doctor or other health worker to lower your blood cholesterol level?**

1 = yes

2 = no

3 = uncertain

8 if *HICH* = 2

9 = insufficient data

CHRX: **Are you taking (in the last two weeks) pills or other medicine prescribed by a doctor to lower your blood cholesterol level?**

1 = yes

2 = no

3 = uncertain

8 if *HICH* = 2

9 = insufficient data

CHRECD: **Have you had your blood cholesterol measured in the last year?**

1 = yes

2 = no

9 = insufficient data

4.1.2.2 Data quality

To ensure that data sent to and received by MONICA Data Centre (MDC) was correct several data checking procedures were performed before data was accepted for the common collaborative database. Even though each centre used their own local questionnaires, they transferred data to MONICA data transfer format before sending it to the MDC. When data was received by MDC, it was checked for constraint violations (systematic errors). The checking of the constraint violations included checking of used codes for each data item and

accepting only those listed in the MONICA data transfer formats as well as some crosschecking between data items for consistency of coding. For example, if sex was 1 (men) there should not be any answers for female questions about the menopausal status and the use of oral contraceptives. [159]

If received data included any constraint violations they were reported back to the originating centre for correction or verification. Reported constraint violations remained in the database as open questions for that centre until they were either corrected or confirmed.

In addition to the checking of the constraint violations in the data, retrospective quality assessment reports were prepared for each data item [10, 13-16, 151-156]. These reports retrospectively assessed how surveys actually were conducted in each centre and reported all deviations from the standard MONICA procedures. Based on these findings, the quality of data for each data item was quantified by quality scores. These quality scores were based on different criteria for different data items. They rated the data quality from 0 to 2. Zero being worst and 2 being perfect.

The response rates of the surveys varied considerably between the populations but also in time within populations. The lowest response rates were below 50% and highest above 90%. [9]

4.2 Simulations

Simulations were used to study the effect of non-response on the cross-sectional and trend estimates (study I), and the effect of the pre-analytic sources of variation on the total cholesterol distribution (study III).

When studying the effect of non-response on the cross-sectional and trend estimates, two independent data sets, 2000 observations each were simulated. For the first dataset, representing the first survey, 43.8% of the observations were assigned to be smokers, and for the second dataset, representing the second survey, 38.4% of the observations were assigned to be smokers. These two simulated datasets were then used to study the changes in the estimates of the

prevalence of daily smoking in three different cases of non-response: missing completely at random, not missing at random but the smoking pattern between respondents and non-respondents remains unchanged between the two surveys, and not missing at random but the smoking pattern between respondents and non-respondents changes between the two surveys.

When studying the effect of the pre-analytic sources of variation on the total cholesterol distribution three differing population total cholesterol distributions were selected from the WHO MONICA Project. From the literature, the magnitudes of the effects of the different pre-analytic methods on the total cholesterol measurement were obtained. Applying the corrections based on the different pre-analytic methods to those three cholesterol distributions, the hypothetical changes to the cholesterol distribution were studied.

4.3 Definition of used indicators

In the studies, different indicators were calculated from the data. For the study I, the level of education was classified into three categories: low, middle, and high. The low educational level includes only primary school or less, the middle educational level includes secondary school or intermediate school between secondary school and university (e.g. technical training) and the high educational level includes university or college education.

For the study II, the body mass index (kg/m^2) was calculated from weight and height measurements.

For study III, the prevalence of hypercholesterolaemia was calculated. Two different definitions were used, total cholesterol $\geq$ 6.5 mmol/l and total cholesterol $\geq$ 5.0 mmol/l.

For study IV, no specific indicators were defined.

For study V, indicators for the prevalence of hypercholesterolaemia, the prevalence of awareness of hypercholesterolaemia, the prevalence of treatment, the prevalence of controlled hypercholesterolaemia, and the prevalence of total cholesterol measurement were calculated.

The prevalence of hypercholesterolaemia is a proportion of survey respondents with:

- Total cholesterol ≥ 6.5 mmol/l.
- Total cholesterol ≥ 6.5 mmol/l or using lipid lowering drugs.
- Total cholesterol ≥ 6.2 mmol/l.
- Total cholesterol ≥ 6.2 mmol/l or using lipid lowering drugs.
- Total cholesterol ≥ 5.0 mmol/l.
- Total cholesterol ≥ 5.0 mmol/l or using lipid lowering drugs.

The prevalence of awareness of hypercholesterolaemia is the proportion of survey respondents who have been told by a doctor or other health professional to have elevated blood cholesterol.

The prevalence of treatment in the entire population is the proportion of survey respondents who have been using lipid-lowering drugs during the past two weeks. When the prevalence refers to those who are aware of their elevated total cholesterol, the prevalence is a proportion of survey respondents aware of their hypercholesterolaemia who have been using lipid-lowering drugs during the past two weeks.

The prevalence of control hypercholesterolaemia is a proportion of survey respondents who were using lipid-lowering drugs and who had total cholesterol < 6.5 mmol/l.

The prevalence of total cholesterol measurement is a proportion of survey respondents who have had their total cholesterol measurement during the past year.

4.4 Statistical methods

Results were presented for age group of 35-64 years for men and women separately, unless otherwise specified. When presenting the results from a wide age range, it is important to use

age standardization. With age standardization, results between populations with different age distributions can be compared.

The direct age standardization [160] with World standard population weights [161] was used when calculating prevalences and means. The weights for each ten year age group in age range 35-64 years are given in *Table 11*.

Table 11. World Standard Population weights

	35-44	**45-54**	**55-64**
Weight	12/31	11/31	8/31

Even though most of the MONICA populations are in Europe, the World standard population weights were chosen since the age distributions in most of the populations, also the European ones, followed quite closely the World population age distribution. [162]

For all statistical analyses, the SAS 8.2 was used.

5 RESULTS

5.1 Effect of response rate on the survey results

Obtaining high response rates is becoming more and more difficult. Survey organizers have been worried about decreasing response rates and their effect on survey results. (Study I)

5.1.1 Changes in response rates

In the MONICA surveys, which were conducted between the early 1980's and the late 1990's, the decline in the response rate could already be seen in most of the population. In the initial surveys, the response rate was on average 73% for both men and women. The response rate declined during the ten years of the study, so that in the final survey, the response rate was on average 69% for men and 72% for women. On average, the decline in response rate between the surveys was 3.8 percentage points in men and 1.0 percentage points in women, but there was wide variation between the populations (*Figure 20*). In both men and women, the response rate declined over 10 percentage points in three populations, the decline being as big as 27 percentage points in one of them.

There was also a clear pattern of change in the response rates by age groups. The response rate tended to increase by the age, i.e. the response rate was lowest among the youngest age groups.

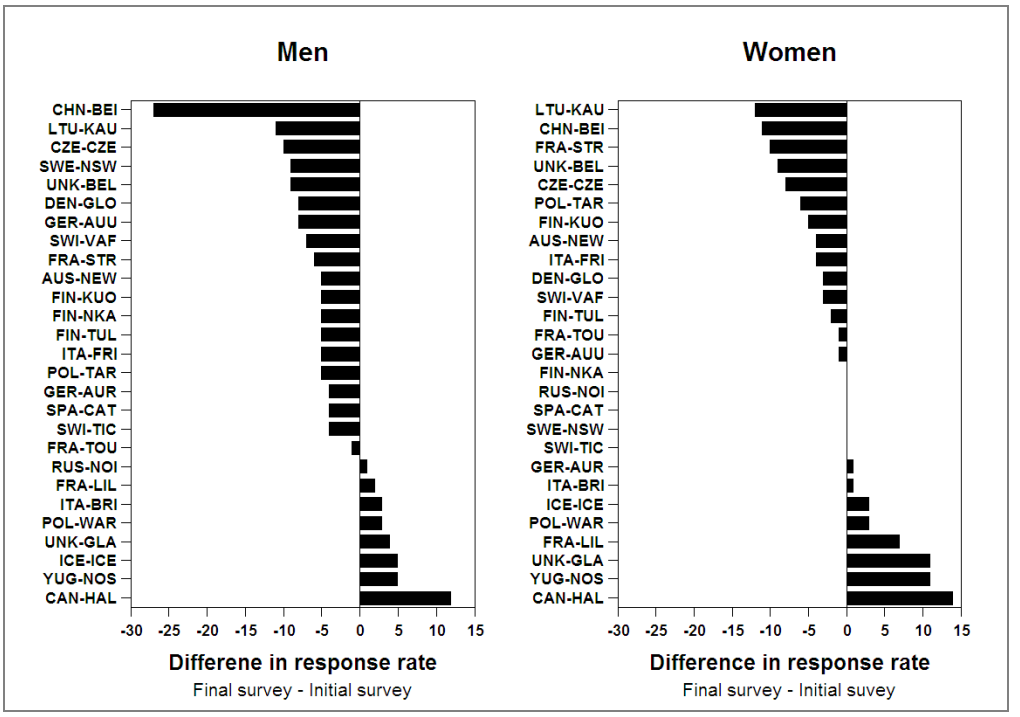

Figure 20. Change in the response rate between the initial and final surveys

5.1.2 Difference between respondents and non-respondents

If the respondents and non-respondents would have a similar socio-economic and health profile, there would not be any concern about the accuracy of the population estimates. Unfortunate this was not the case. Across the populations, non-respondents tended to be more often unmarried, and had a lower education level than respondents. The non-respondents also were more often daily smokers and in many populations were more frequently using drug treatment for their hypertension. On the other hand, the non-respondents also tended to be leaner than the respondents.

5.1.3 Effect of non-response on population estimates and trends

These observed differences between the respondents and the non-respondents were likely to bias the results that were based on the respondents only. The results from the respondents tended to overestimate the proportion of being married or cohabiting. The respondent data also underestimated the proportion of having low education and among men, overestimated the proportion of being highly educated. Among women, respondent data did not show a clear pattern of over- or underestimation of being highly educated. The results from respondents tended to underestimate the population prevalence of smoking as well as the prevalence of hypertension drug treatment. The non-respondents did not affect the BMI results.

The decline in response rate between the surveys affected also the trend estimates of daily smoking. In men, the smoking trend declined in most populations in both respondents and non-respondents, but the decline was slower among the non-respondents. In six populations, the smoking trend declined in the respondents but increased in the non-respondents. In women, the pattern was not as clear-cut. In half of the populations, the smoking trend increased in both the respondents and the non-respondents, but the increase was faster among the non-respondents. In the other half, there were populations, where the smoking trend decreased in both the respondents and the non-respondents, and populations where the smoking trend increased in the respondents but decreased in the non-respondents. These differing trends between the respondents and the non-respondents caused the bias to the smoking trends when they were based only on the respondents. In men, the respondents tended to overestimate the smoking trend while in women the trend was overestimated in half of the populations and underestimated in the other half.

The simulation results verified these finding. When the response rate declines the precision of the cross-sectional and trend estimates also declines. If the respondents and non-respondents are alike, the estimates are not affected by the declining response rate but as soon as the respondents and non-respondents differ, the declining response rate will affect the accuracy of the estimate.

5.1.4 Quality implications

Declining response rates were shown to have a significant effect on results in the WHO MONICA Project. These finding were also supported by simulation studies, which showed that if non-response is not random, the estimates are already affected when response rate drops to 80%. It should be noted that by increasing the sample size to compensate for non-response, the problem is just moved from the low precision to the higher precision of the wrong estimates. The importance of high response rate gets emphasized in the trend analysis.

5.2 Using the quantified data quality indicators

Data quality is a rather complicated phenomenon being a combination of representativeness of the sample and accuracy of the measurements with several subcomponents. In the WHO MONICA Project, a system of quality scores was developed to help the evaluation of the data quality. These quality scores are artificial quantifiers of the data quality, which are based on pre-defined criteria. (Study II)

5.2.1 Effect of data quality on the results

Quality scores were used in the analysis to test the robustness of the results in relation to the data quality. Initially the analysis included all available data regardless their quality. To study the effect of data quality on the results, the analysis was repeated but now with adjustment for the data quality. This adjustment was made in two ways, first by removing the populations with quality problems and secondly, by weighting the analysis by the quality scores.

When studying the correlation between the mean total cholesterol and the mean BMI in the WHO MONICA Project final risk factor surveys, the results changed significantly after the adjustment for the data quality. The correlation coefficient for men was 0.64 ($p = 0.00$) and for women 0.06 ($p = 0.72$) when all the available data were used regardless of their quality. Removing the populations with quality problems the correlation coefficient for men was 0.44

(p = 0.06) and for women –0.14 (p = 0.57). Weighting of the analysis by the data quality did not have as drastic effect. The correlation coefficient for men was 0.63 (p = 0.00) and for women 0.05 (p = 0.76).

5.2.2 Quality implications

Quality of data was shown to have a significant effect on the results. Conclusions drawn from the complete data regardless of the data quality may results in misleading interpretations. Sensitivity analysis incorporating the data quality in the analysis will increase the reliability of the results and reduce the risk for wrong conclusions.

5.3 Effect of measurement procedures on the survey results

Different measurement procedures are known to affect the outcome. The effect of season, posture of the subject, prolonged use of tourniquet during the blood drawing, and the type of blood sample (plasma/serum) on total cholesterol measurement was studied in three different populations (*Figure 21*). The first population (A) had low total cholesterol level with mean of 4.0 mmol/l (SD 1.01), and the majority of the population had total cholesterol below 5.0 mmol/l (94 %). The second population (B) had an average total cholesterol level with mean of 5.4 mmol/l (SD 1.05), and about half (46 %) of the population had total cholesterol below 5.0 mmol/l. The third population (C) had high total cholesterol level with mean of 6.2 mmol/l (SD 0.95), and only 7 % of the population had total cholesterol below 5.0 mmol/l. (Study III)

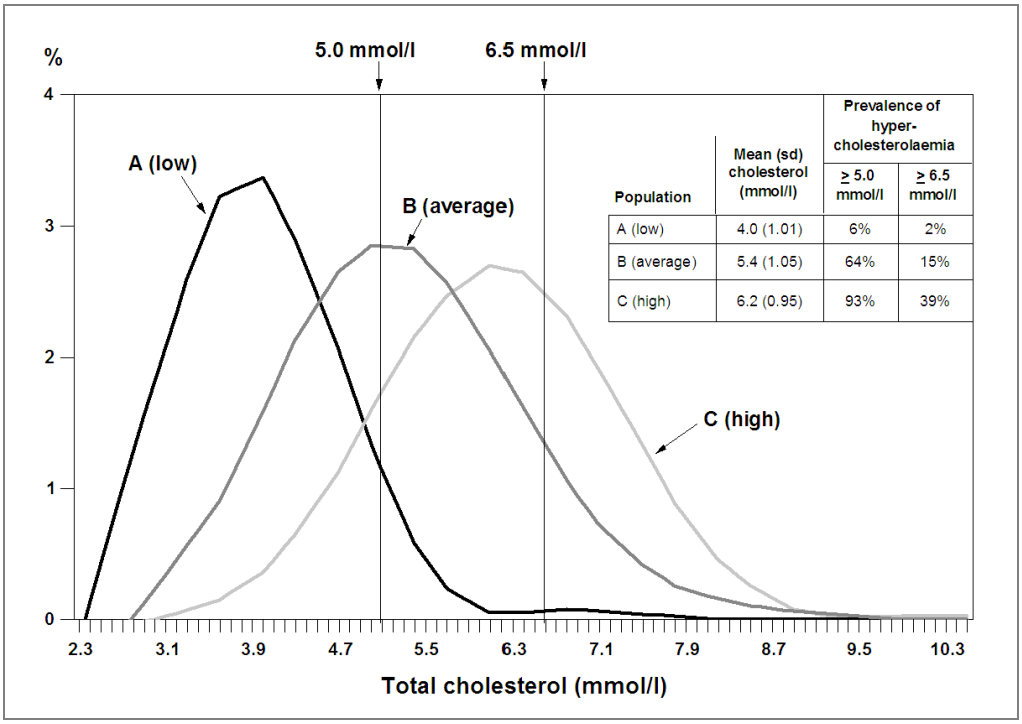

Figure 21. Total cholesterol distribution in populations A (low), B (average), and C (high)

5.3.1 Effect on the prevalence of hypercholesterolaemia

In population A, the change of season for blood drawing would change the prevalence of hypercholesterolaemia by 3 % when cut-point of 5.0 mmol/l is used. The prevalence of hypercholesterolaemia with definition ≥ 6.5 mmol/l does not change, because the majority of the population will remain below that cut-point. (*Figure 22*)

The change in posture during the blood drawing has a marked effect on the prevalence of hypercholesterolaemia (≥ 5.0 mmol/l). When changing from supine to sitting, the prevalence may increase by 6 %. The change to standing posture would increase the prevalence even more (14 %) but blood samples are rarely ever drawn in standing posture. (*Figure 22*)

The prolonged use of tourniquet could increase the prevalence of hypercholesterolaemia (≥ 5.0 mmol/l) by 3 % and the type of blood sample may change the prevalence from 1-4 %. (*Figure 22*)

The changes in estimates caused by the change in procedure are generally small, except for change in posture. This is so because the major part of the population has very low total cholesterol levels, and effects on procedures will not increase the levels so much that they would cross the cut-points of 5.0 mmol/l or 6.5 mmol/l.

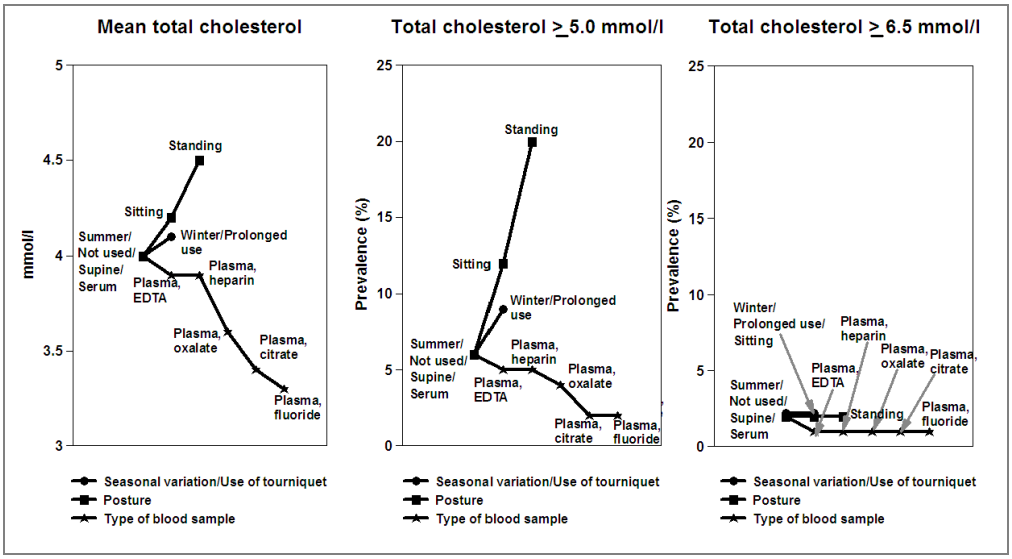

Figure 22. Effect of pre-analytic procedures on population A (low)

In population B, the change of season may have a 3 % effect on the prevalence of hypercholesterolaemia with cut-point of 5.0 mmol/l and a 2 % effect with cut-point of 6.5 mmol/l. The change in posture during the blood drawing from supine to sitting could increase the prevalence by 8 % with cut-point 5.0 mmol/l and by 4 % with cut-point 6.5 mmol/l. The effect of prolonged tourniquet use on the prevalence of hypercholesterolaemia may be upto 3 % with cut-point 5.0 mmol/l and 2 % with cut-point 6.5 mmol/l. The type of blood sample from which the total cholesterol is determined has a significant effect on the prevalence of hypercholesterolaemia. When 5.0 mmol/l cut-point is used, the effect may be 5-36 % and with

cut-point 6.5 mmol/l, 2-13 %. The effects are big as most of the population has total cholesterol around the cut-points and even with small changes will easily shift from one side of the cut point to the other. (*Figure 23*)

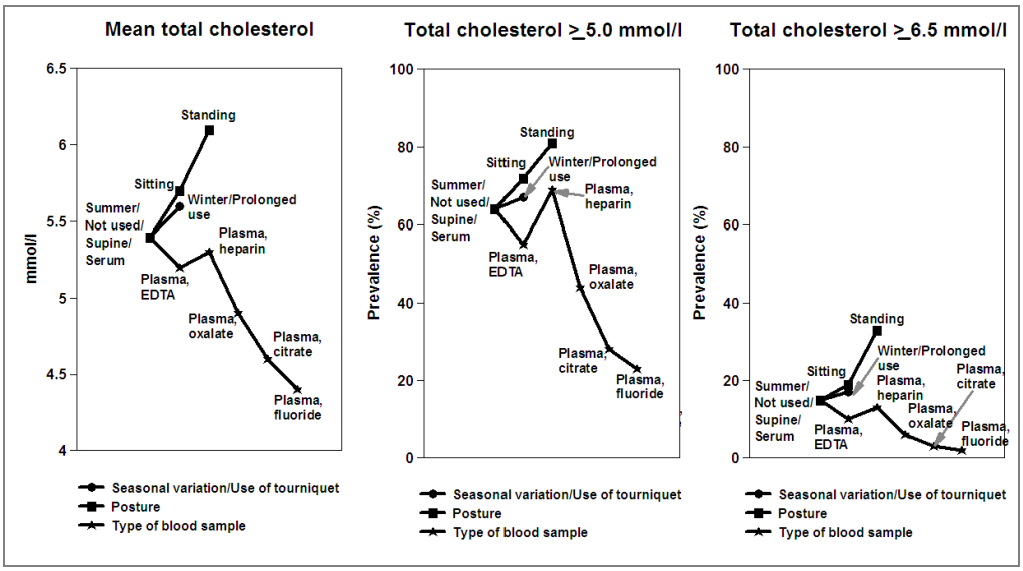

Figure 23. Effect of pre-analytic procedures on population B (average)

In population C, the change of season may have a 1 % effect on the prevalence of hypercholesterolaemia with 5.0 mmol/l cut-point and a 7 % effect with 6.5 mmol/l cut-point. The change in posture from supine to sitting may change the prevalence by 3 % when the cut-point is 5.0 mmol/l and by 13 % when the cut-point is 6.5 mmol/l. The prolonged used of a tourniquet during the blood drawing could increase the prevalence by 1 % if a cut-point 5.0 mmol/l is used, and by 7 % if a cut-point 6.5 mmol/l is used. The effect of the type of blood sample on the prevalence varied from 3 % to 37 % with cut-point 5.0 mmol/l and from 3 % to 36 % with cut-point 6.5 mmol/l. (*Figure 24*)

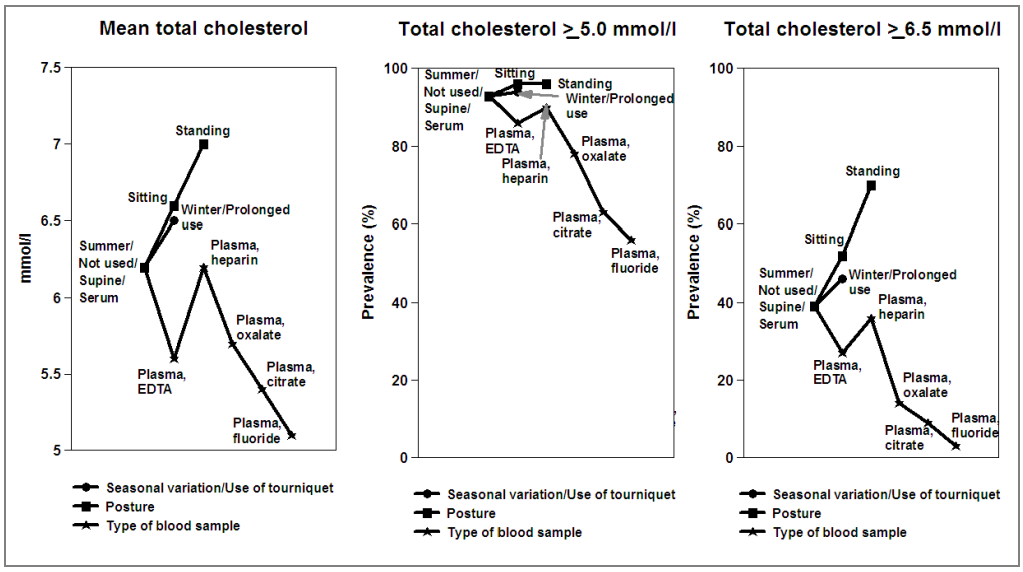

Figure 24. Effect of pre-analytic procedures on population C (high)

In the WHO MONICA Project, a standardized protocol was used also for the pre-analytic procedures of the total cholesterol measurements. There were few deviations from the given protocol. The effect of these deviations on total cholesterol levels was estimated. Even though the total cholesterol level in some populations would change slightly after correcting for the used pre-analytic procedures, the ranking of the populations based on their total cholesterol levels remains essentially the same as before corrections.

5.3.2 Quality implications

It was shown that the change of the measurement procedures might alter the results markedly. When comparing results between studies, it is important to note that observed differences, if small, may be due to different measurement procedures. When surveys are conducted regularly to monitor the population level changes in total cholesterol or any other indicator, the measurement procedures should be kept unchanged between surveys to ensure the accuracy and reliability of the trend estimates.

5.4 Assessing the data quality

The retrospective quality assessment for questionnaire items on awareness and treatment of high cholesterol was made to evaluate the availability and quality of these data item. These questions were introduced to the MONICA protocol in 1990 and by that time the initial survey was already conducted in all populations and in many populations the middle survey was underway. Therefore, the data for awareness and treatment of high cholesterol is available only for some populations from the middle surveys and for most populations from the final survey. (Study IV)

5.4.1 Quality of data on awareness of high cholesterol

The question about the awareness of high cholesterol (*"Have you ever been told by a doctor or other health worker that you have high blood cholesterol?"*) was available from 32% of the populations in the middle survey and from 95% of the populations in the final survey. In the middle survey the question differed from the standard MONICA questions in most of the populations but it was considered that data could be extracted from them without major problems. In the final survey, most of the populations used the exact wording of the MONICA question. In all cases where the local question deviated from the MONICA question, the difference was that phrase *"by a doctor or other health worker"* was omitted.

Table 12. Quality scores for the awareness of high cholesterol

	Quality score			**N**
	2 (no concern)	**1 (some concern)**	**0 (major concern)**	
Middle survey	7 (64%)	3 (27%)	1 (9%)	11
Final survey	28 (80%)	0 (0%)	7 (20%)	35

The availability of data among respondents was generally good in both the middle and final survey. The quality score for the awareness of high cholesterol was 2 (no concern) in 64% of the populations in the middle survey and in 80% of the populations in the final survey (*Table 12*).

5.4.2 Quality of data on dietary treatment of high cholesterol

The question about the dietary treatment of high cholesterol (*"Are you on special diet prescribed by a doctor or other health worker to lower your blood cholesterol level?"*) was available from 20% of the populations in the middle survey and from 90% of the population in the final survey. In the middle survey, most of the populations used local questions, which deviated from the MONICA question. The deviation was considered to be minor in all except one case, where the data for MONICA couldn't be extracted from the local question. In the final survey, most of the populations used the MONICA question as it was given in the MONICA Manual. The biggest deviation between the local and MONICA questions was the omission of the phrase *"prescribed by a doctor or other health worker"*.

The availability of the data among the respondents was in general good in both the middle and final survey, with only a few exceptions in the final survey, where there were over 30% of missing data. The quality score for the dietary treatment of high cholesterol was 2 (no concern) in 64% of the populations in the middle survey and in 66% of the populations in the final survey (*Table 13*).

Table 13. Quality scores for the dietary treatment of high cholesterol

	Quality score			N
	2 (no concern)	**1 (some concern)**	**0 (major concern)**	
Middle survey	7 (64%)	0 (0%)	4 (36%)	11
Final survey	23 (66%)	2 (6%)	10 (29%)	35

5.4.3 Quality of data on drug treatment of high cholesterol

The question about the drug treatment of high cholesterol (*"Are you taking (in the last two weeks) pills or other medicine prescribed by a doctor to lower your blood cholesterol level?"*) was available from 37% of the populations in the middle survey and from 100% of the populations in the final survey. In the middle survey, most of the populations had used local questions, which deviated from the MONICA question. Also in the final survey, a little over half of the populations used questions different from MONICA question. The most common deviation between local and MONICA questions was the omission of the phrase *"prescribed by a doctor"*.

The availability of the data among the respondents was generally good in both the middle and final surveys, except in a few populations in the final survey, where the proportion of missing data was over 30%. The quality score for the drug treatment of high cholesterol was 2 (no concern) in 54% of the populations in the middle survey and in 72% of the populations in the final survey (*Table 14*).

Table 14. Quality scores for the drug treatment of high cholesterol

	Quality score			N
	2 (no concern)	**1 (some concern)**	**0 (major concern)**	
Middle survey	7 (54%)	0 (0%)	6 (46%)	13
Final survey	26 (72%)	3 (8%)	7 (19%)	36

5.4.4 Quality of data on cholesterol measurement during the last year

The question about the cholesterol measurement during the last year (*"Have you had your blood cholesterol measured in the last year?"*) was available from 24% of the populations in the middle survey and from 90% of the populations in the final survey. In the middle survey, most of the populations had local question different from the MONICA question. In the final

survey, the majority of the populations used the exact wording of the MONICA question. In general, the questions, which deviated from the MONICA question asked *"When was your cholesterol last measured?"*.

Table 15. Quality scores for the cholesterol measurement in the last year

	Quality score			N
	2 (no concern)	**1 (some concern)**	**0 (major concern)**	
Middle survey	6 (50%)	3 (25%)	3 (25%)	12
Final survey	27 (77%)	2 (6%)	6 (17%)	35

The availability of the data among the respondents was generally good in both middle and final surveys, except in one population in the final survey, where the proportion of missing data was over 30%. The quality score for the cholesterol measurement during the last year was 2 (no concern) in 50% of the populations in the middle survey and in 77% of the populations in the final survey (*Table 15*).

5.4.5 Quality implications

It was shown that even in a study like the WHO MONICA Project, where a lot of effort was put into the standardization of the methods, there are deviations between the surveys. When these deviations in the methods are big enough or the data availability among the respondents is low, the quality of the available data is questionable. Assessment of the data quality will point out possible shortcomings of the surveys and when these shortcomings are known, they can be taken into account during the analysis and interpretations of the results.

5.5 Reporting the survey results

Reporting of the survey results is usually done using different kinds of indicators derived from the collected data. These indicators can be based on one variable or a combination of several variables. The definition of the indicators has an effect on the results. These effects of indicator definition were studied using the data from the total cholesterol measurements and the drug treatment of high cholesterol in the WHO MONICA Project final risk factor surveys. (Study V)

5.5.1 Effect of selected indicators

Six different definitions of the prevalence of hypercholesterolaemia (see *Chapter 4.3*) were used to see how sensitive the prevalence of hypercholesterolaemia was to the indicator definition in the MONICA populations. When comparing the prevalence of hypercholesterolaemia defined as total cholesterol ≥ 5.0 mmol/l with the prevalence using the definition total cholesterol ≥ 6.5 mmol/l, the prevalence changed on average 49 percentage points in both men and women. In men, the change was between 22 percentage points and 58 percentage points. In women, the change ranged from 24 percentage points to 57 percentage points. (*Figure 25*)

The comparison of the prevalence of hypercholesterolaemia once defined as total cholesterol ≥ 5.0 mmol/l and secondly as total cholesterol ≥ 6.2 mmol/l resulted in slightly smaller differences than above. For men, the difference was on average 41 percentage points, the minimum being 21 percentage points and the maximum being 50 percentage points. For women, the difference was on average 41 percentage points, with the minimum being 23 percentage points and the maximum being 48 percentage points. (*Figure 25*)

When the prevalence definitions total cholesterol ≥ 6.2 mmol/l and total cholesterol ≥ 6.5 mmol/l were compared, the differences decreased significantly as could be expected. In men, the difference was on average nine percentage points, the minimum being zero percentage

points and the maximum being 14 percentage points. In women, the difference was on average eight percentage points, the minimum being one percentage point and the maximum being 13 percentage points. (*Figure 25*)

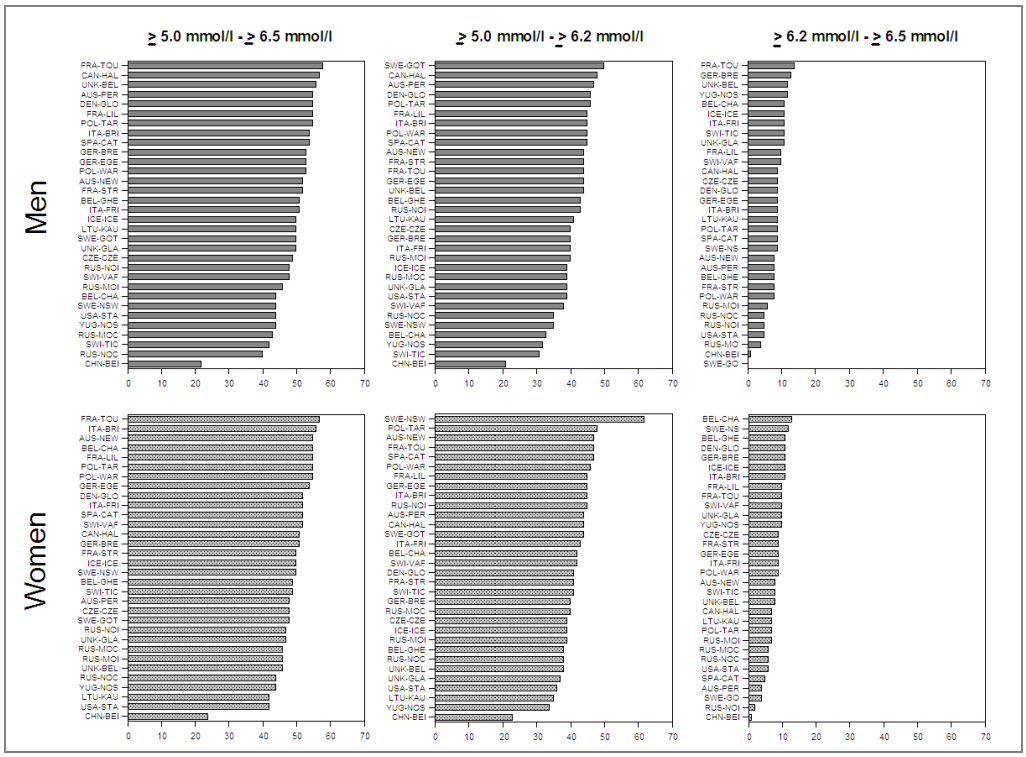

Figure 25. The difference in the prevalence of hypercholesterolaemia between two different definitions

If the use of lipid lowering drugs was taken into account when defining the prevalence of hypercholesterolaemia, the prevalence increased on average by one percentage point in all three cut-points.

5.5.2 Effect of survey year

The final surveys were conducted between the years 1990 and 1996. It could be expected that the prevalence of hypercholesterolaemia and its treatment were higher in the surveys conducted around the mid-1990's than in surveys conducted at the beginning of the 1990's. This difference is due to increased awareness of high cholesterol as a risk factor for cardiovascular diseases and the decreased cost of measurements and medications.

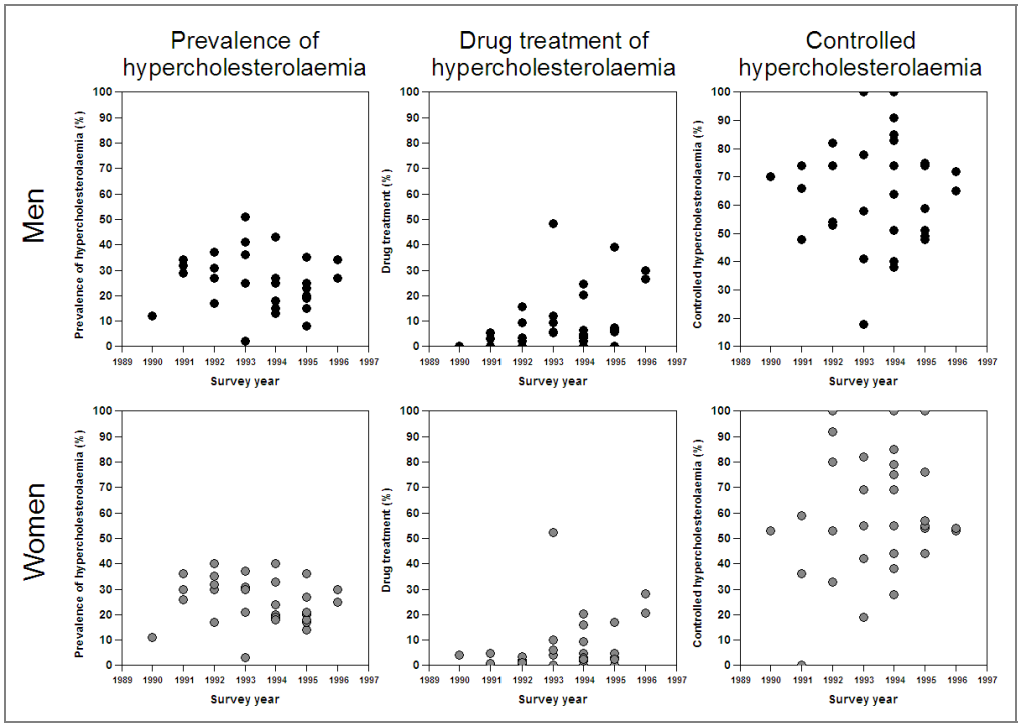

Figure 26. The effect of survey year on the prevalence of hypercholesterolaemia, drug treatment of hypercholesterolaemia, and the controlled hypercholesterolaemia

There was no relationship between the prevalence of hypercholesterolaemia (total cholesterol ≥ 6.5 mmol/l) and the year of the survey, but the drug treatment of hypercholesterolaemia seemed to be higher in surveys conducted towards the mid-1990's in comparison to surveys conducted at the beginning of the 1990's. The prevalence of controlled hypercholesterolaemia

did not increase or decrease over time, i.e. the increase in the drug treatment over time did not mean that more hypercholesterolaemics would have their condition under control by the drug treatment. (*Figure 26*)

The year of the survey did not affect the prevalence of total cholesterol measurements. The prevalence was equal between the surveys conducted in early 1990's and in the mid-1990's. Also, the awareness of hypercholesterolaemia was not affected by the survey year. (*Figure 27*)

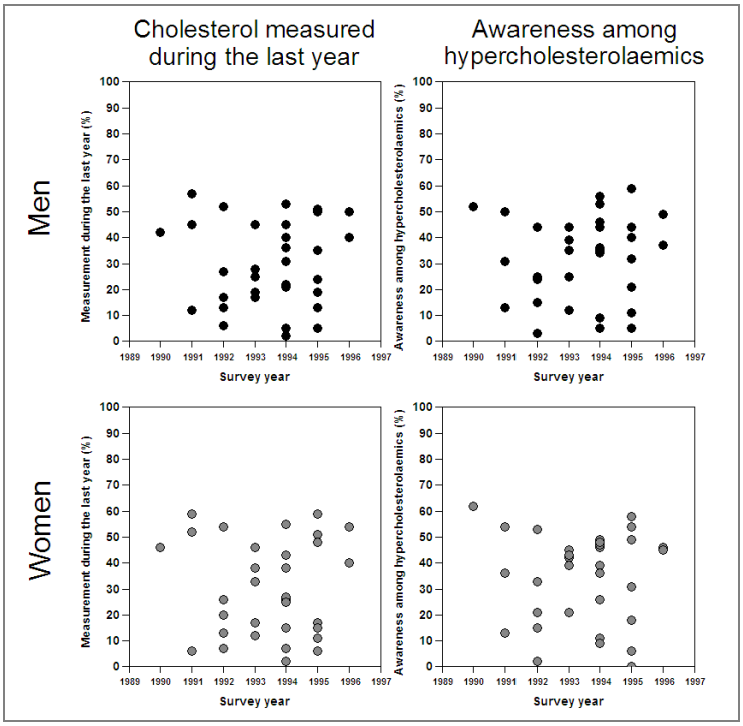

Figure 27. Effect of survey year on the prevalence of cholesterol measurement during the last year, and the awareness of high cholesterol among hypercholesterolaemics

5.5.3 Quality implications

As was shown, the definition of the indicators used makes a big difference in the results. This is important to recognise when comparing the results between surveys. As many health indicators change over time, it is important to make sure that the surveys are conducted at the same time (year) when cross-sectional comparisons are made. Otherwise, the results are not directly comparable due to change in phenomena over time.

6 SUMMARY DISCUSSION

6.1 Strengths and limitations of the study

This thesis was mostly based on the data from the WHO MONICA Project. These data provided an excellent foundation for studying the effects of data quality on the survey results, as the data collection was based on well-standardized protocol, which allowed the comparisons between over 30 populations in 21 countries on four continents. The data also were subjected to detailed checking for systematic data-entry errors and the retrospective quality assessment. This quality checking provided detailed documentation about the achieved data quality.

In the different study populations, the response rates and the achieved data quality varied considerably. Generally, from an epidemiological point of few, this would be a limitation of the study since low response rates and low data quality would bias the results and make the interpretation of data more difficult. In this thesis, the variety of response rates and the differing data quality was strength because it made it possible to compare the effects of different response rates and data quality on the results of the health surveys.

In the WHO MONICA Project, some data were also collected from the non-respondents using a standardized questionnaire. This unique set of non-respondent data made it possible to study how much the results based only respondents are biased, when compared with the results from both respondents and non-respondents. The limitation was the low availability of non-respondent data, which caused uncertainty of the results.

There were several aspects of total survey error, which could not be studied using the data from the WHO MONICA Project, due to lack of required information or due to MONICA study design. For example, the effect of incentives on the response rates could not be studied, as there was no information about possibly used incentives.

6.2 Importance of the results

The results have shown that data quality has an important role for the accurate and reliable inference from the results of the health surveys. It is not just good coverage of the target population, i.e. the representative sample and high response rate, which makes results accurate and reliable. Even with good coverage, the results can be inaccurate and unreliable if there is measurement error or high item non-response rate.

It was shown that when the response rates go below 80%, the accuracy and the precision of the population estimate starts to decrease. The limit of 70% for response rate given generally in survey literature [2, 25, 74] will in many cases give biased results. The decreasing response rate between the surveys should also be taken seriously, as it has a significant effect on the trend estimates. In some cases, the observed change in the trend estimates could be solely due to the decline in the response rates.

As non-respondents are not evenly distributed though the population groups but tend to concentrate in young, unmarried men with low education, the importance of high response rate is emphasized. The increasing absence of these people from surveys will increase the bias of the results. When the non-response can be identified as being concentrated in certain population groups, the survey organizers should think about ways how to attract these people to the survey. Are these young men too busy to participate? Do they not like paper questionnaires? These are just two examples of questions to be asked and the answer to them may help find ways to increase the participation of young, unmarried men with low education.

Incorporating information about data quality into the analysis helps to validate the robustness of results with regard to observed data quality. This assessment of sensitivity of the results to data quality increases the reliability of the results.

The effects of pre-analytic measurement procedures on total cholesterol results demonstrated the importance of standardization of the measurement protocols. The standardization of methods is important when results are compared with previous surveys from the same

population or between populations. If the measurement procedures are different, the inference of the results will not be reliable, since the observed changes and differences may be due to the differences in measurement procedures.

When the results between the surveys are compared, the correctness of comparisons depends on the accuracy and reliability of data, but also on the comparability of indicators used. It was demonstrated by the prevalence of hypercholesterolaemia, that population level result can be very sensitive for the changes in the indicator definition.

All results presented here provide strong support for the standardization of survey methods and for proper quality control. With standardization and quality control, the accuracy, reliability and comparability of the results can be improved.

6.3 Challenges for future surveys

The increasing knowledge of survey methods, the development of measurement procedures, and the changes in legislation are creating new challenges for future surveys. In many countries, the legislation on personal privacy is getting stricter, which often means increased difficulties in getting permissions to draw representative samples from population registers or other sampling frames. Also, a person's right to limit the use of their information in registers may cause problems for survey sampling.

The changes in legislation may also have an effect on the measurements taken during the survey. An example is the blood pressure measurement, which generally is made using a mercury sphygmomanometer. Due to toxicity of mercury, many countries have already banned the use of mercury sphygmomanometers and the EU is going to ban the use of mercury within the next few years [163, 164]. The alternatives for the mercury sphygmomanometers are automated blood pressure measurement devices, which have been on the market for a few decades but still are quite unreliable [165, 166]. Therefore, the big challenge is to find reliable blood pressure measurement devices to replace the mercury

sphygmomanometers so that the trend estimates and the comparisons between the populations will not be affected by the change in device.

There has also been development toward less expensive measurements, especially among biochemical measurements, which usually require laboratories for the determination. The use of the laboratories is expensive; the measurements have to go through internal and preferably also external quality control to be reliable and transferring samples to the laboratory and storing them at different stages requires well-organized logistics. New techniques for the measurement of cholesterol have been developed. These so-called dry chemistry kits are easy to use in the field and will not require sending of the samples to the laboratory. There is room for improvement of the accuracy of these dry chemistry devices [167-169] and therefore, each device adopted for the use in health surveys should be validated.

The development and changes of the measurement devices will be one of the big challenges for future surveys, since these changes may have an effect on the trends. It is crucial to make sure that new devices get properly validated and pass these validations before they are accepted for the use in surveys.

For questionnaires, the challenges are more regarding the administration mode than the actual questions, which can easily be kept unchanged over time. The increasing number of mobile telephones and the decrease in fixed-line telephones will create new challenges for telephone interviews [49, 59]. Earlier, when there was one telephone line per house, the telephone surveys were traditional household surveys. Now when people are using more mobile phones and are giving up their fixed-line telephones, the nature of telephone surveys is changing. Each member of a household may have his/her own mobile phone and in some cases, people have more than one mobile phone. This change in structure of telephone ownership will also change the nature of telephone surveys from household surveys to personal surveys.

The Internet also has to be considered as a questionnaire administration mode. At the moment, the coverage of the Internet access is not wide enough for a representative population sample and also sampling frames are missing, but in future, this will change. The

Internet may be one answer to the problem of getting young men to participate in health surveys. For many young men, the Internet is much easier to use than regular pen and paper.

Health surveys have been conducted for decades but results between them are in many cases not comparable due to different methods used for the data collection, or due to low response rates, different age range covered by the sample etc. Even when the primary goal is to compare the survey results with previous results from the same population, and the methods and target population are kept unchanged, the declining response rates may jeopardize this goal. People are showing survey fatigue, since they face all kinds of marketing surveys, polls, etc. and the health survey is just another one of those surveys. Making people feel that it is important to participate in a specific health survey is a difficult task and requires new innovative ideas.

When the goal of the survey is to collect data and compare results with other surveys from other populations, standardization and quality control are key concerns. If the data are not collected using standardized methods and no quality control is done, the comparability of the results is impossible to judge. During the years, there has been only one international project, the WHO MONICA Project [7], which has developed a standardized study protocol and conducted quality control to ensure the comparability of results between the populations. A few other international projects with standardized study protocol exist (for example, CINDI) or have existed (Seven countries study) but they are lacking the uniform quality control system to ensure that the standardized protocol is actually used and used correctly.

The WHO has several initiatives on the standardization of health surveys on an international level. The World Health Survey [170] is a questionnaire based survey, where standardized questionnaires [171] are used to collect health information in different populations. The STEPwise approach to risk factors (STEPS) [172] is a survey protocol, which provides three steps: Step 1 – questionnaire based, Step 2 – questionnaire and physical measurements, and Step 3 – questionnaires, physical measurements and biochemical measurements. The STEPS has developed a standardized questionnaire but also the protocol for the physical and biochemical measurements to be used in different populations. In the European context, the

EU has worked on the standardization of survey methods through Health Monitoring Programme [173].

7 CONCLUSIONS

Health surveys are needed for the monitoring of populations' health status, diseases and their determinants, and changes in them. Based on this trend information, preventive activities can be planned and evaluated. Such information is extremely valuable for health policy, health interventions, and health systems. In order to achieve that, the information collected by health surveys has to be accurate and reliable. This can be ensured by the use of standardized survey methods and quality control.

There is still a lot of misunderstanding and ignorance about the benefits of standardization and quality control. It is often said to be too expensive to use highly standardized methods and to conduct good quality control. Actually, it can be a waste of money to conduct a survey without proper standardization and quality control, and to obtain data, which are unreliable and cannot be used for anything.

In the WHO and EU, some development has been under way to prepare standardized guidelines for the health survey organizers. For these initiatives, proper quality control component are essential.

International recommendations on quality are extremely important for countries, which are in the process of starting health surveys and have no prior experience. However, it is not enough to have international recommendations. There should also be enough skilful people available to help survey organizers in the adoption of the recommendations for the local use and in the training of the survey personnel to use those recommendations correctly.

REFERENCES

1. Last JM. *A Dictionary of Epidemiology*. 4th ed. New York: Oxford University Press; 2001.

2. Lessler JT, Kalsbeek WD. *Nonsampling error in surveys*. New York: John Wiley & Sons, Inc.; 1992.

3. Warwick DP, Lininger CA. *The sample survey: theory and practice*. New York: McGraw-Hill, Inc.; 1975.

4. Rothman KJ, Greenland S. *Modern epidemiology*. 2nd ed. Philadelphia: Lippincott-Raven Publishers; 1998.

5. Biemer PP, Lyberg LE. *Introduction to survey quality*. Hoboken: John Wiley & Sons, Inc.; 2003.

6. Tunstall-Pedoe H, for the WHO MONICA Project. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators. *J Clin Epidemiol* 1988;**41**(2):105-14.

7. Tunstall-Pedoe H, for the WHO MONICA Project, editors. *MONICA Monograph and multimedia sourcebook. World's largest study of heart disease, stroke, risk factors, and population trends, 1979-2002*. Geneva: World Health Organization; 2003.

8. Kuulasmaa K, Tolonen H, Ferrario M, Ruokokoski E, for the WHO MONICA Project *Age, date of examination and survey periods in the MONICA surveys* (1998) Available from: URL:http://www.ktl.fi/publications/monica/age/ageqa.htm, URN:NBN:fi-fe19991075

9. Wolf HK, Kuulasmaa K, Tolonen H, Ruokokoski E, for the WHO MONICA Project *Participation rates, quality of sampling frames and sampling fractions in the MONICA surveys* (1998) Available from: URL:http://www.ktl.fi/publications/monica/nonres/nonres.htm, URN:NBN:fi-fe19991076

10. Kuulasmaa K, Hense H, Tolonen H, for the WHO MONICA Project *Quality assessment of data on blood pressure in the WHO MONICA Project* (1998) Available from: URL:http://www.ktl.fi/publications/monica/bp/bpqa.htm, URN:NBN:fi-fe19991082

11. Mähönen M, Tolonen H, Kuulasmaa K, for the WHO MONICA Project *Quality assessment of stroke event registration data in the WHO MONICA Project* (1998) Available from: URL:http://www.ktl.fi/publications/monica/strokeqa/strokeqa.htm, URN:NBN:fi-fe19991080

12. Mähönen M, Tolonen H, Kuulasmaa K, Tunstall-Pedoe H, Amouyel P, for the WHO MONICA Project *Quality assessment of coronary event registration data in the WHO MONICA Project* (1999) Available from: URL:http://www.ktl.fi/publications/monica/coreqa/coreqa.htm, URN:NBN:fi-fe19991072

13. Molarius A, Kuulasmaa K, Evans A, McCrum E, Tolonen H, for the WHO MONICA Project *Quality assessment of data on smoking behaviour in the WHO MONICA Project* (1999) Available from: URL:http://www.ktl.fi/publications/monica/smoking/qa30.htm, URN:NBN:fi-fe19991077

14. Lundberg V, Tolonen H, Kuulasmaa K, for the WHO MONICA Project *Quality assessment of data on menopausal status and hormones in the WHO MONICA Project* (1999) Available from: URL:http://www.ktl.fi/publications/monica/womenqa/womenqa.htm, URN:NBN:fi-fe19991131

15. Tolonen H, Ferrario M, Minoja M, for the WHO MONICA Project *Quality assessment of data on awareness and treatment of high cholesterol in the WHO MONICA Project* (1999) Available from: URL:http://www.ktl.fi/publications/monica/hich/hchdrug.htm, URN:NBN:fi-fe19991130

16. Tolonen H, Kuulasmaa K, for the WHO MONICA Project *Quality assessment of data on use of aspirin in the WHO MONICA Project* (1999) Available from: URL:http://www.ktl.fi/publications/monica/aspirin/aspirinqa.htm, URN:NBN:fi-fe19991129

17. Tolonen H, Kuulasmaa K, Laatikainen T, Wolf HK, and the European Health Risk Monitoring Project *Recommendation for indicators, international collaboration, protocol and*

manual of operations for chronic disease risk factor surveys. (2002) Available from: URL:http://www.ktl.fi/publications/ehrm/product2/title.htm, URN:NBN:fi-fe20021443

18. Aday AL. *Designing and conducting health surveys*. San Francisco: Jossey-Bass Publishers; 1996.

19. Dillman DA. *Mail and Telephone Surveys. The total design method*. New York: John Wiley & Sons, Inc.; 1978.

20. Dillman DA. *Mail and internet surveys: the tailored design methods*. 2nd ed. New York: John Wiley & Sons, Inc.; 1999.

21. Fowler FJJ. *Improving survey questions: design and evaluation*. Thousand Oaks: Sage Publications, Inc.; 1995.

22. Groves R. *Telephone survey methodology*. New York: John Wiley & Sons, Inc.; 1988.

23. Groves RM, Fowler FJJ, Couper MP, Lepkowski JM, Singer E, Tourangeau R. *Survey Methodology*. New Jersey: John Wiley & Sons, Inc.; 2004.

24. Harkness JA, van de Vijver FJR, Ph Mohler P, editors. *Cross-cultural survey methods*. New York: John Wiley & Sons, Inc.; 2002.

25. Salant P, Dillman DA. *How to conduct your own survey*. New York: John Wiley & Sons, Inc.; 1994.

26. Tourangeau R, Rips LJ, Rasinski K. *The psychology of survey response*. London: Cambridge University Press; 2000.

27. Rose GA, Blackburn H, Gillium RF, Prineas RJ. *Cardiovascular Survey Methods*. 2nd ed. Geneva: World Health Organization; 1982.

28. WHO MONICA Project *MONICA Manual, Part III: Population survey, Section 1: Population survey data component* (1997) Available from: URL:http://www.ktl.fi/publications/monica/manual/part3/iii-1.htm, URN:NBN:fi-fe19981151

29. WHO *STEPS Instruments for NCD risk factors (Core and expanded version 1.4)* (2003) Available from: URL:http://www.who.int/ncd_surveillance/en/stepsinstrumentcore_exp_v1.4.pdf,

30. WHO *STEPS Field manual. Guidelines for field staff* (2003) Available from: URL:http://www.who.int/ncd_surveillance/steps/riskfactor/en/fieldmanualL_july.pdf,

31. WHO *STEPS Field Manual. Appendices* (2003) Available from: URL:http://www.who.int/ncd_surveillance/steps/riskfactor/en/fieldmanual_appendicesfeb04.pdf,

32. WHO *STEPS Planning and implementation* (2004) Available from: URL:http://www.who.int/ncd_surveillance/steps/en/planningmanualaugust03.pdf,

33. Fisher G, Pappas G, Limb M. Prospects, problems, and prerequisites for national health examination surveys in developing countries. *Soc Sci Med* 1996;**42**(12):1639-50.

34. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bull World Health Organ* 2001;**79**(4):373-4.

35. Luepker RV, Evans A, McKeigue P, Reddy KS. *Cardiovascular Survey Methods*. 3rd ed. Geneva: World Health Organization; 2004.

36. National Library of Medicine *PubMed ages* (2004) Available from: URL:http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Limits&DB=pubmed,

37. *International Statistical Classification of Diseases and Related Health Problems. Tenth Revision. Volume 2. Instruction manual*. Geneva: World Health Organization; 1993.

38. Redfern P. Population registers: some administrative and statistical pros and cons. *J R Stat Soc Ser A Stat Soc* 1989;**152**(1):1-41.

39. Population Register Centre *Population Information System* (2004) Available from: URL:http://www.vaestorekisterikeskus.fi/vrk/home.nsf/maindocuments/15f53d1df5b83c3dc2256eee003fdb6a?opendocument,

40. National Tex Board *Population registration in Sweden* (2000) Available from: URL:http://www.rsv.se/broschyrer/711b/711b03.pdf,

41. AS Andmevara *Population Register* (2004) Available from: URL:http://www.andmevara.ee/andmek/Popreg.htm,

42. Office of Citizenship and Migration Affairs *Functions* (2003) Available from: URL:http://www.ocma.gov.lv/?_p=189&menu__id=3,

43. Kuzmickas B *Republic of Lithuania. Law on the population register* (1992) Available from: URL:http://www.litlex.lt/Litlex/Eng/Frames/Laws/Documents/44.HTM,

44. Statistics Netherlands *Figures - Population counter* (2003) Available from: URL:http://www.cbs.nl/en/figures/popclock/popclocknl.asp,

45. U.S. Census Bureau. Census 2000 Basics. Washington DC: U.S. Government Printing Office; 2002.

46. National Statistics *What is the Census?* (2002) Available from: URL:http://www.statistics.gov.uk/census2001/cb_1.asp,

47. Statistics Finland. Population Census 2002. Handbook. Helsinki: Statistics Finland; 2001.

48. Barnes R. Use of registers and population censuses as sampling frames in Great Britain. *Stat J UN Econ Comm Eur* 1986;**4**(1):81-92.

49. Kuusela V, Simpanen M. Effects of mobile phones on telephone survey. Practices and results. In: ICIS Conference; 2002; Copenhagen, Denmark; 2002.

50. National Statistics *Living in Britain - General Household Survey 2002* (2003) Available from: URL:http://www.statistics.gov.uk/downloads/theme_compendia/lib2002.pdf,

51. Federal Statistical Office Germany *Household budget surveys and time use* (2003) Available from: URL:http://www.destatis.de/basis/e/evs/bidgtab6.htm,

52. Behavioral Risk Factor Surveillance System *2000 telephone coverage* (2003) Available from: URL:http://www.cdc.gov/brfss/technical_infodata/surveydata/2000/table3_00.htm,

53. International Telecommunication Union *Basic indicators: Population, GDP, total telephone subscribers and total telephone substribers per 100 people* (2004) Available from: URL:http://www.itu.int/ITU-D/ict/statistics/at_glance/basic03.pdf,

54. International Telecommunication Union *Main telephone lines, subscribers per 100 people* (2004) Available from: URL:http://www.itu.int/ITU-D/ict/statistics/at_glance/main03.pdf,

55. International Telecommunication Union *Mobile cellular, subsribers per 100 people* (2004) Available from: URL:http://www.itu.int/ITU-D/ict/statistics/at_glance/cellular03.pdf,

56. Anderson JE, Nelson DE, Wilson RW. Telephone coverage and measurement of health risk indicators: data from the National Health Interview Survey. *Am J Public Health* 1998;**88**(9):1392-5.

57. Adimora AA, Schoenbach VJ, Martinson FE, Stancil TR, Donaldson KH. Driver's license and voter registration lists as population-based sampling frames for rural African Americans. *Ann Epidemiol* 2001;**11**(6):385-8.

58. Collins M, Sykes W. The Problems of Non-Coverage and Unlisted Numbers in Telephone Surveys in Britain. *Journal of the Royal Statistical Society Series a-Statistics in Society* 1987;**150**:241-53.

59. Kuusela V. Mobile Phones and Telephone Survey Methods. In: The ASC's 4th International Conference on Survey and Statistical Computing; 2003; Warwick University, UK; 2003.

60. Beerten R, Martin J. Household ownership of telephone and other communication links: implications for telephone surveys. *Survey Methodological Bulletin* 1999;**44**:1-7.

61. Kuusela V. Telephone coverage in Finland. *International BLAISE User Group Newsletter* 1998;**no 11**.

62. O'Mahony PG, Thomson RG, Rodgers H, Dobson R, James OF. Accuracy of the family health services authority register in Newcastle upon Tyne, UK as a sampling frame for population studies. *J Epidemiol Community Health* 1997;**51**(2):206-7.

63. *The election commission* (2004) Available from: URL:http://www.electoralcommission.org.uk/,

64. *Register to vote in your state by using this postcard form and guide* (2003) Available from: URL:http://www.fec.gov/votregis/pdf/nvra.pdf,

65. Bowling A, Hart D, Silman A. Accuracy of electoral registers and Family Practitioner Committee lists for population studies of the very elderly. *J Epidemiol Community Health* 1989;**43**(4):391-4.

66. Jamieson A, Shin HB, Day J. Voting and registration in the election of November 2000. U.S. Census Bureau; 2002.

67. Lynch CF, Logsden-Sackett N, Edwards SL, Cantor KP. The driver's license list as a population-based sampling frame in Iowa. *Am J Public Health* 1994;**84**(3):469-72.

68. Kuulasmaa K. Sampling. In: Tunstall-Pedoe H, for the WHO MONICA Project, editors. MONICA Monograph and Multimedia Sourcebook. World's largest study of heart disease, stroke, risk factors, and population trends, 1979-2002. Geneva: World Health Organization; 2003.

69. Cochran WG. *Sampling techniques*. 3rd ed. New York: John Wiley & Sons, Inc.; 1977.

70. Lemeshow S, Hosmer DW, Klar J, Lwanga SK. *Adequacy of sample size in health studies*. Chichester: John Wiley & Sons, Inc.; 1990.

71. Thompson SK. *Sampling*. 2nd ed. New York: John Wiley & Sons, Inc.; 2002.

72. Levy PS, Lemeshow S. *Sampling of populations: methods and applications*. 3rd ed. New York: John Wiley & Sons, Inc.; 1999.

73. Rao PSRS. *Sampling methodologies with applications*. New York: Chapman & Hall/CRC; 2000.

74. Rea LM, Parker RA. *Designing and conducting survey research. A comprehensive guide*. San Francisco: Jossey-Bass; 1997.

75. Fowler FJJ, Mangione TW. *Standardized survey interviewing. Minimizing interviewer-related error*. 2nd ed. Newbury Part: Sage Publications; 1990.

76. Okamoto K, Ohsuka K, Shiraishi T, Hukazawa E, Wakasugi S, Furuta K. Comparability of epidemiological information between self- and interviewer-administered questionnaires. *J Clin Epidemiol* 2002;**55**(5):505-11.

77. Belson WA. *The design and understanding of survey questions*. Aldershot: Grower; 1981.

78. Converse JM, Presser S. *Survey questions. Handcrafting the standardized questionnaire*. Newbury Park: Sage Publications; 1986.

79. Sudman S, Bradburn NM. *Asking questions. A practical guide to questionnaire design*. San Fancisco: Jossey-Bass; 1982.

80. Sudman S, Bradburn NM, Schwarz N. *Thinking about answers: the application of congnitive processes to survey methodology*. San Francisco: Jossey-Bass; 1996.

81. Swain L. Basic principles of questionnaire design. *Survey Methods* 1985;**11**(2):161-70.

82. Tanur J. *Questions about questions. Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation; 1992.

83. Suessbrick A, Schober MF, Conrad FG. Different respondents interpret ordinary questions quite differently. Proceedings of the Section on Survey Research Methods. American Statistical Association; 2000:907-12.

84. Biemer PP, Groves R, Lyberg LE, Mathiowez N, Sudman S, editors. *Measurement errors in surveys*. New York: John Wiley & Sons, Inc.; 1991.

85. Apodaca R, Lea S, Edwards B. The effect of longitudinal burden on survey participation. Proceedings of the Section on Survey Research Methods. American Statistical Association; 1998:906-10.

86. Treat JB. The effect of questionnaire length on response. Proceedings of the Section on Survey Research Methods. American Statistical Association; 1996:734-39.

87. Herzog AR, Bachman JG. Effect of questionnaire length on response quality. *Public Opionion Quarterly* 1982;**45**:549-59.

88. Oksenberg L, Cannell C, Kalton G. New Strategies for Pretesting Survey Questions. *Journal of Official Statistics* 1991;**7**(3):346-65.

89. Presser S, Blair J. Survey pretesting: Do different methods produce different results? *Sociological Methodology* 1994;**24**:73-104.

90. Carmines EG, Zeller RA. *Reliability and validity assessment*. Beverly Hills: Sage Publications; 1979.

91. DeVellis RF. *Scale development. Theory and applications*. 2nd ed. Thousand Oaks: Sage Publications; 2003.

92. *Designing to meet the needs of visually impaired people* (2004) Available from: URL:http://www.textmatters.com/tm/guides/visually_impaired.html,

93. Hartley J. *Designing intructional text*. 3rd ed. New Jersey: Nichols Publishing Company; 1994.

94. Raggett D, Le Hors A, Jacobs I, editors *HTML 4.01 Specification* (1999) Available from: URL:http://www.w3.org/TR/html4/,

95. Sun Mircosystems I *Products & Technologies. Java Technology* (2004) Available from: URL:http://java.sun.com/,

96. Chisholm W, Vanderheiden G, Jacobs I, eds. *List of checkpoints for Web content accessibility guidelines 1.0* (1999) Available from: URL:http://www.w3.org/TR/WAI-WEBCONTENT/checkpoint-list.html,

97. Mangione TW. *Mail Surveys. Improving the Quality*. Thousand Oaks: Sage Publications, Inc.; 1995.

98. Aoki K, Elasmar M. Opportunities and challenges of conducting web surveys: results of a field experiment. Proceedings of the Section on Survey Research Methods. American Statistical Association; 2000:930-35.

99. International Telecommunication Union *Internet indicators: Hosts, Users and number of PCs* (2004) Available from: URL:http://www.itu.int/ITU-D/ict/statistics/at_glance/Internet03.pdf,

100. U.S. Census Bureau. Home computers and internet use in the United States: August 2000. Washington DC: U.S. Governmental Printing; 2001.

101. *Nordic Information Society Statistics 2002*. Helsinki: Yliopistopaino; 2002.

102. Ramirez C, Sharp K, Forster L. Mode effects in an internet/paper survey of employees. Proceedings of the Section on Survey Research Methods. American Statistical Association; 2000:1033-38.

103. Kwak N, Radler B. A comparison between mail and web surveys: response pattern, respondent profile, and data quality. *Journal of Official Statistics* 2002;**18**(2):257-73.

104. Radon K, Goldberg M, Becklake M, Pindur U, Hege I, Nowak D. Low acceptance of an internet-based online questionnaire by young adults. *Epidemiology* 2002;**13**(6):748-9.

105. Tourangeau R, Couper MP. Cognitive issues in the design of web surveys. Proceedings of the Section on Survey Research Methods. American Statistical Association; 2000:476-81.

106. Mickey RM, Vacek PM, Worden JK. Effects of survey mode and advance letters on contact and interview completion rates for population-based surveys of women. Proceedings of the Section on Survey Research Methods. American Statistical Association; 1999:854-58.

107. Greenfield TK, Midanik LT, Rogers JD. Effects of telephone versus face-to-face interview modes on reports of alcohol consumption. *Addiction* 2000;**95**(2):277-84.

108. Mishra SI, Dooley D, Catalano R, Serxner S. Telephone health surveys: potential bias from noncompletion. *Am J Public Health* 1993;**83**(1):94-9.

109. Barnett S, Franks P. Telephone ownership and deaf people: implications for telephone surveys. *Am J Public Health* 1999;**89**(11):1754-6.

110. Sykes WM, Collins M. Comparing telephone and face-to-face interviewing in the United Kingdom. *Survey Methodology* 1987;**13**(1):15-28.

111. Weeks MF, Kulka RA, Lessler JT, Whitmore RW. Personal versus telephone surveys for collecting household health data at the local level. *Am J Public Health* 1983;**73**(12):1389-94.

112. Cannell CF, Oksenberg L, Converse JM. Experiments in interviewing techniques. Ann Arbor: Institute for Social Research; 1979.

113. Bason JJ. Comparison of telephone, mail, web, and IVR surveys of drug and alcohol use among university of Georgia students. Proceedings of the Section on Survey Research Methods. American Statistical Association; 2000:1074-80.

114. Battistutta D, Byth K, Norton R, Rose G. Response rates: a comparison of mail, telephone and personal interview strategies for an Australian population. *Community Health Stud* 1983;**7**(3):309-13.

115. Siemiatycki J. A comparison of mail, telephone, and home interview strategies for household health surveys. *Am J Public Health* 1979;**69**(3):238-45.

116. Siemiatycki J, Campbell S, Richardson L, Aubert D. Quality of response in different population groups in mail and telephone surveys. *Am J Epidemiol* 1984;**120**(2):302-14.

117. Brogger J, Bakke P, Eide GE, Gulsvik A. Comparison of telephone and postal survey modes on respiratory symptoms and risk factors. *Am J Epidemiol* 2002;**155**(6):572-6.

118. Elliott DB, Fowler FJJ. Response order effects in the medicare population: the interaction between mode of survey administration and respondent age. Proceedings of the Section on Survey Research Methods. American Statistical Association; 2000:936-40.

119. van Ooijen M, Ivens UI, Johansen C, Skov T. Comparison of a self-administered questionnaire and a telephone interview of 146 Danish waste collectors. *Am J Ind Med* 1997;**31**(5):653-8.

120. Weir P, Beri S, O'Colmain B. A comparison and evaluation of two survey data collection methodologies: CATI vs. mail. Proceedings of the Section on Survey Research Methods. American Statistical Association; 1999:388-93.

121. National Center for Health Statistics *1999-Current National Health and Nutrition Examination Survey (NHANES). Survey Operations Manuals, Brochures, Consent Documents* (2004) Available from: URL:http://www.cdc.gov/nchs/about/major/nhanes/currentnhanes.htm,

122. Erens B, Primatesta P, Prior G, Eds. *Volume 1: Findings, Volume 2: Methodology & Documentation* (2001) Available from: URL:http://www.archive.official-documents.co.uk/document/doh/survey99/hse99.htm,

123. Tolonen H, Wolf HK, Jakovljevic D, Kuulasmaa K, for the European Health Risk Monitoring Project *Review of surveys for risk factors of major chronic diseases and comparability of the results* (2002) Available from: URL:http://www.ktl.fi/publications/ehrm/product1/title.htm, URN:NBN:fi-fe20021442

124. WHO MONICA Project *MONICA Manual* (1999) Available from: URL:http://www.ktl.fi/publications/monica/manual/index.htm, URN:NBN:fi-fe19981146

125. Schober MF, Conrad FG, Fricker SS. When and how should survey interviewers clarify question meaning? Proceedings of the Section on Survey Research Methods. American Statistical Association; 1999:985-90.

126. Williams A, Cook T, Apodaca R, Hall K. Retaining good interviewers: characteristics that affect interviewer turnover and retention in a CAPI survey. Proceedings of the Section on Survey Research Methods. American Statistical Association; 1999:1032-35.

127. de Heer WF, Israëls AZ. Response trends in Europe. Proceedings of the Section on Survey Research Methods. American Statistical Association; 1992:92-101.

128. Brennan M, Hoek J. The behavior of respondents, nonrespondents, and refusers across mail surveys. *Public Opionion Quarterly* 1992;**56**:530-35.

129. Shahar E, Folsom AR, Jackson R. The effect of nonresponse on prevalence estimates for a referent population: insights from a population-based cohort study. Atherosclerosis Risk in Communities (ARIC) Study Investigators. *Ann Epidemiol* 1996;**6**(6):498-506.

130. Groves RM, Couper MP. Correlates of nonresponse in personal visit surveys. Proceedings of the Section on Survey Research Methods. American Statistical Association; 1992:102-11.

131. Forthofer RN. Investigation of nonresponse bias in NHANES II. *Am J Epidemiol* 1983;**117**(4):507-15.

132. Voigt LF, Koepsell TD, Daling JR. Characteristics of telephone survey respondents according to willingness to participate. *Am J Epidemiol* 2003;**157**(1):66-73.

133. Chapman DD. Reporting complete nonresponse in a sample survey. Proceedings of the Section on Survey Research Methods. American Statistical Association; 1991:266-71.

134. Savitz DA. *Interpreting epidemiological evidence. Strategies for study desing and analysis*. New York: Oxford University Press; 2003.

135. Cottler LB, Zipp JF, Robins LN, Spitznagel EL. Difficult-to-recruit respondents and their effect on prevalence estimates in an epidemiologic survey. *Am J Epidemiol* 1987;**125**(2):329-39.

136. Strandberg TE, Salomaa VV, Vanhanen HT, Naukkarinen VA, Sarna SJ, Miettinen TA. Mortality in participants and non-participants of a multifactorial prevention study of cardiovascular diseases: a 28 year follow up of the Helsinki Businessmen Study. *Br Heart J* 1995;**74**(4):449-54.

137. Korkeila K, Suominen S, Ahvenainen J, Ojanlatva A, Rautava P, Helenius H, et al. Non-response and related factors in a nation-wide health survey. *Eur J Epidemiol* 2001;**17**(11):991-9.

138. Jackson R, Chambless LE, Yang K, Byrne T, Watson R, Folsom A, et al. Differences between respondents and nonrespondents in a multicenter community-based study vary by

gender ethnicity. The Atherosclerosis Risk in Communities (ARIC) Study Investigators. *J Clin Epidemiol* 1996;**49**(12):1441-46.

139. Criqui MH, Barrett-Connor E, Austin M. Differences between respondents and non-respondents in a population-based cardiovascular disease study. *Am J Epidemiol* 1978;**108**(5):367-72.

140. Bond GG, Lipps TE, Stafford BA, Cook RR. A comparison of cause-specific mortality among participants and nonparticipants in a work-site medical surveillance program. *J Occup Med* 1991;**33**(6):677-80.

141. Bergstrand R, Vedin A, Wilhelmsson C, Wilhelmsen L. Bias due to non-participation and heterogenous sub-groups in population surveys. *J Chronic Dis* 1983;**36**(10):725-8.

142. Eaker S, Bergstrom R, Bergstrom A, Adami HO, Nyren O. Response rate to mailed epidemiologic questionnaires: a population-based randomized trial of variations in design and mailing routines. *Am J Epidemiol* 1998;**147**(1):74-82.

143. Jones J. The effects of non-response on statistical inference. *J Health Soc Policy* 1996;**8**(1):49-62.

144. Chun YI, Robertson KW. The effect of advance letters and reminder/thank you letters on reducing nonrespondense in an establisment survey: an experimental study. Proceedings of the Section on Survey Research Methods. American Statistical Association; 1995:1151-56.

145. Fox RJ, Crask MR, Kim J. Mail survey response rate. A meta-analysis of selected techniques for inducing response. *Public Opionion Quarterly* 1988;**52**:467-91.

146. Spry VM, Hovell MF, Sallis JG, Hofsteter CR, Elder JP, Molgaard CA. Recruiting survey respondents to mailed surveys: controlled trials of incentives and prompts. *Am J Epidemiol* 1989;**130**(1):166-72.

147. Bergsten JW, Weeks MF, Bryan FA. Effects of an advance telephone call in a personal interview survey. *Public Opionion Quarterly* 1984;**48**:650-57.

148. Allison KR, Yoshida KK. Increasing response rates in community health surveys administered by telephone. *Can J Public Health* 1989;**80**(1):67-70.

149. Harlow BL, Crea EC, East MA, Oleson B, Fraer CJ, Cramer DW. Telephone answering machines: the influence of leaving messages on telephone interviewing response rates. *Epidemiology* 1993;**4**(4):380-3.

150. Schweitzer M, Asch DA. Timing payments to subjects of mail surveys: cost-effectiveness and bias. *J Clin Epidemiol* 1995;**48**(11):1325-9.

151. Molarius A, Kuulasmaa K, Sans S, for the WHO MONICA Project *Quality assessment of weight and height measurements in the WHO MONICA Project* (1998) Available from: URL:http://www.ktl.fi/publications/monica/bmi/bmiqa20.htm, URN:NBN:fi-fe19991079

152. Ferrario M, Kuulasmaa K, Grafnetter D, Moltchanov V, for the WHO MONICA Project *Quality assessment of total cholesterol measurement in the WHO MONICA Project* (1999) Available from: URL:http://www.ktl.fi/publications/monica/tchol/tcholqa.htm, URN:NBN:fi-fe19991083

153. Marques-Vidal P, Ferrario M, Kuulasmaa K, Grafnetter D, Moltchanov V, for the WHO MONICA Project *Quality assessment of data on HDL cholesterol in the WHO MONICA Project* (1999) Available from: URL:http://www.ktl.fi/publications/monica/hdl/hdlqa.htm, URN:NBN:fi-fe19991137

154. Molarius A, Sans S, Kuulasmaa K, for the WHO MONICA Project *Quality assessment of data on waist and hip circumferences in the WHO MONICA Project* (1998) Available from: URL:http://www.ktl.fi/publications/monica/waisthip/waisthipqa.htm, URN:NBN:fi-fe19991091

155. Molarius A, Kuulasmaa K, Moltchanov V, Ferrario M, for the WHO MONICA Project *Quality assessment of data on marital status and educational achievement in the WHO MONICA Project* (1998) Available from: URL:http://www.ktl.fi/publications/monica/educ/educqa.htm, URN:NBN:fi-fe19991078

156. Molarius A, Tuomilehto J, Kuulasmaa K, for the WHO MONICA Project *Quality assessment of data on hypertension control in the WHO MONICA Project* (1998) Available from: URL:http://www.ktl.fi/publications/monica/hyperten/hbpdrug.htm, URN:NBN:fi-fe19991084

157. Tolonen H, Mähönen M, Kuulasmaa K, Nissinen A *Reporting the risk factor survey data* (2002) Available from: URL:http://www.ktl.fi/publications/ehrm/product3/title.htm, URN:NBN:fi-fe20031091

158. WHO MONICA Project *MONICA Manual, Part III: Population survey, Section 2: Standardization of lipid measurements* (1998) Available from: URL:http://www.ktl.fi/publications/monica/manual/part3/iii-2.htm, URN:NBN:fi-fe19981152

159. WHO MONICA Project *MONICA Manual, Part V: Data transfer and analysis* (1998) Available from: URL:http://www.ktl.fi/publications/monica/manual/part5/v-1.htm, URN:NBN:fi-fe19981160

160. Armitage P, Berry G. *Statistical methods in medical research*. Oxford: Blackwell Scientific Publications; 1987.

161. Waterhouse J, Muir CS, Correa P, Powell J, editors. *Cancer incidence in five continents*. Lyon: IARC; 1976.

162. Tolonen H. Age standardization. In: Tunstall-Pedoe H, for the WHO MONICA Project, editors. MONICA Monograph and multimedia sourcebook. World's largest study of heart disease, stroke, risk factors, and population trends, 1979-2002. Geneva: World Health Organization; 2003.

163. O'Brien E. Will mercury manometers soon be obsolete? *J Hum Hypertens* 1995;**9**(12):933-4.

164. Markandu ND, Whitcher F, Arnold A, Carney C. The mercury sphygmomanometer should be abandoned before it is proscribed. *J Hum Hypertens* 2000;**14**(1):31-6.

165. O'Brien E. Replacing the mercury sphygmomanometer. Requires clinicians to demand better automated devices. *BMJ* 2000;**320**(7238):815-6.

166. British Hypertension Society *Validated Blood Pessure Monitors List* (2005) Available from: URL:http://www.bhsoc.org/blood_pressure_list.htm,

167. Rubin DA, McMurray RG, Harrell JS, Carlson BW, Bangdiwala S. Accuracy of three dry-chemistry methods for lipid profiling and risk factor classification. *Int J Sport Nutr Exerc Metab* 2003;**13**(3):358-68.

168. James R, Tyler C, Henrikson D. An evaluation of the accuracy of the Reflotron system in the field. *Med J Aust* 1988;**149**(3):130-1.

169. Kaufman HW, McNamara JR, Anderson KM, Wilson PW, Schaefer EJ. How reliably can compact chemistry analyzers measure lipids? *JAMA* 1990;**263**(9):1245-9.

170. *WHO World Health Survey. Survey Manual* (2002) Available from: URL:http://www3.who.int/whs/P/Survey_Manual.pdf,

171. World Health Survey *New Instrument and Related Docuemtns* (2003) Available from: URL:http://www3.who.int/whs/P/instrumentandrel8293.html,

172. Bonita R, de Courten M, Dwyet T, Jamrozik K, Winkelmann R *Surveillance of risk factors for noncommunicable diseases: The WHO STEPwise approach. Summary* (2001) Available from: URL:http://www.who.int/ncd_surveillance/en/summaryfinal_rev1_english.pdf,

173. European Commission *Programme of Community action in the field of public health (2003-2008)* (2004) Available from: URL:http://europa.eu.int/comm/health/ph_programme/programme_en.htm,