EHES Manual

PART A.
PLANNING AND PREPARATION OF THE SURVEY

2nd edition

Hanna Tolonen (editor)

EHES

Measuring the Health of Europeans

Edited by

Hanna Tolonen

EHES MANUAL

PART A.
PLANNING AND PREPARATION
OF THE SURVEY

2nd edition

Helsinki 2016

Layout: Hanna Tolonen

Cover: Hanna Tolonen

Graphics:

Chapter 3: Johan Heldal and Susie Jentoft

Chapter 10: Georg Alfthan

Chapter 12: Ari Haukijärvi

Other Chapters: Hanna Tolonen

Contents

Introduction

PART A

The European Health Examination Survey (EHES) Manual provides guidelines and specifies the requirements for the implementation of standardized national health examination surveys (HES) in the European countries. Recommendations based on past experiences from national and international surveys were prepared by the Feasibility of a European Health examination Survey (FEHES) Project (Tolonen 2008). The EHES manual builds on these recommendations and on further experience obtained during the EHES Pilot Project in 2009-2012. The EHES Manual has three parts:

A. Planning and preparation of the survey
B. Fieldwork procedures
C. European level coordination

The EHES Manual is maintained by the EHES Reference Centre. This is the 2nd edition of the EHES Manual on which many topics are further clarified, providing more details and examples. The plan is to update it also in future. The latest version of the EHES Manual is available in the Internet at www.ehes.info.

This is Part A of the EHES Manual. It provides guidelines for the planning and preparation of national health examination surveys.

As part of the planning of a national HES, each country has to prepare a national HES Manual. The procedures described in the national manual should follow the European standards specified in the EHES Manual. The national manual should be specific also in issues where the EHES manual can only give alternatives or general guidelines. The EHES manual is unspecific in situations where the national circumstances vary and there is no common procedure which could be reasonably followed in all countries. When the European recommendation differs from the procedure used in earlier national surveys, the procedure to be adapted in the new national HES needs to be considered carefully. Sometimes there may be need to compromise between European comparability and the possibility to follow national trends from the past. The countries should prepare the national manuals in collaboration with the EHES Reference Centre.

References

- Tolonen H, Koponen P, Aromaa A, et al. (Eds.) Recommendations for the Health Examination Surveys in Europe. B21/2008, Publications of the National Public Health Institute, Helsinki 2008. Also available from http://urn.fi/URN:ISBN:978-951-740-838-7

1. Survey planning and management

Päivikki Koponen[1], Hanna Tolonen[1], Arpo Aromaa[1], Kari Kuulasmaa[1]

[1] National Institute for Health and Welfare (THL), Helsinki, Finland

PART A

High quality planning and management are the keys to achieving the survey's objectives. The planning process ensures that the survey can be effectively implemented in a reasonable time, within the budget and with the highest quality that is affordable and consistent with the aims and purposes (Franklin & Walker 2003). All survey plans need to be repeatedly overhauled depending on the progress of action. This requires efficient management. This chapter focuses on national activities in the planning and preparation of the national surveys, in particular on survey management.

1.1 Survey process

The first step in planning and preparation includes defining the aims and purposes of the survey. These will be the basis for selecting the topics and actions of the data collection. They will also guide the decisions on how the EHES standards will be implemented in the national survey. The aims and purposes of the national survey should rely on national and European level health policies, and information needs. National health care systems, previous and current health surveys and expertise available in the country will also affect the feasibility of different options. All decisions need to be made in the context of previous national HISs and HESs, as well as other major health surveys in each country. If there are other national surveys, such as surveys on nutrition, lifestyles or health behaviour, or other health interview surveys, the new HES needs to be timed and tailored to fit in the national health survey system. An evaluation of already exisiting data sorces is needed to define if the HES is the best way to collect the data. As the national HESs are anticipated to be repeated with regular intervals, the survey planning process needs to be ongoing with experiences and results on previous surveys leading to the next phase of data collection (Figure 1.1).
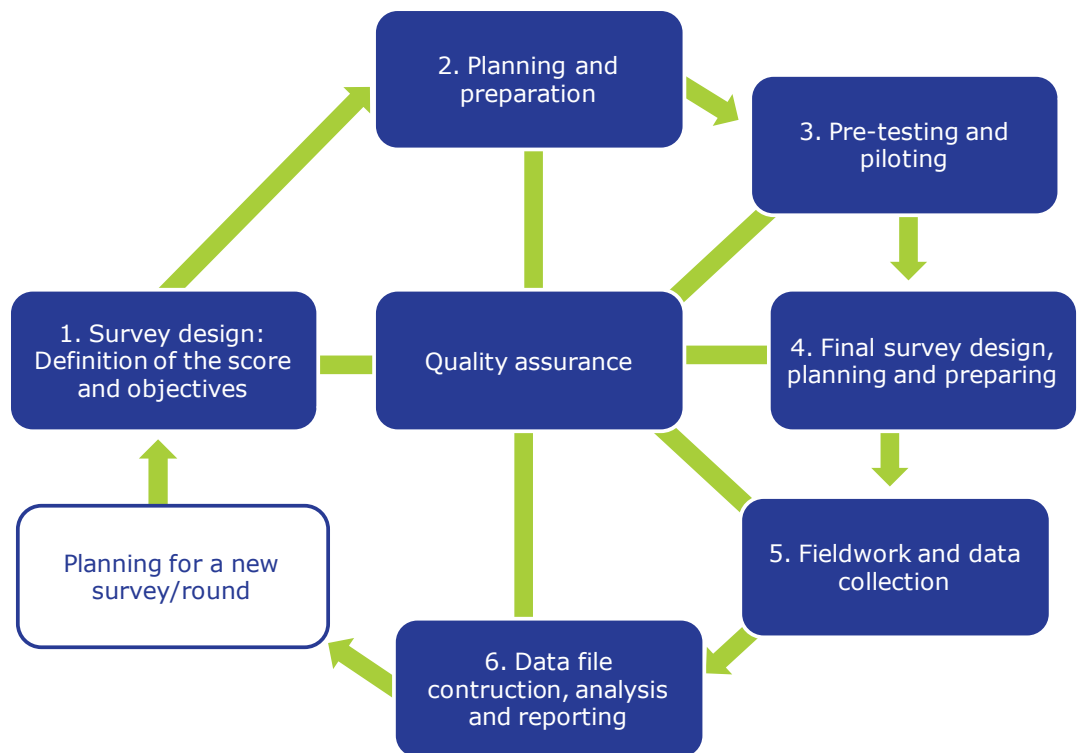
Figure 1.1 Stages in the survey process (stages adapted from Fraklin & Walker 2003, Czaja & Blair 2005)

Six main stages in the survey processa are shown in Figure 1.1. Even though proceed after each other, there is a need to return to previous stages throughout the survey process to adapt the plans according to experiences and feedback from different stakeholders, as well as the feedback from the fieldwork staff and participants.

- The output of stage 1 of the planning of the survey is the first version of the survey proposal. Commitment from key organizations such as the ministry and the national public health institute, national statistical institute and other relevant organizations can be sought based on these preliminary plans and ideas. The survey management structure is also defined at this stage as well as a preliminary time schedule for the survey.

- Stage 2 includes the detailed planning of the sampling, survey contents, fieldwork (e.g. timing) and data collection, data management as well as a preliminary plan for analysis and reporting. Thus the duration of this stage is relatively long - between 6 to 12 months. The output is a detailed survey plan with the budget and a first draft of the survey manual including the questionnaires, and measurement protocols, and other materials (information leaflets, consent forms etc). Ethical approval is sought based on the detailed proposal.

- Stage 3 includes pretesting and piloting. After this, the proposals and manuals, as well as all survey materials (including the computer programs, survey web-sites as well as communication plans) can be finalized.

- Stage 4 sets up the fieldwork and data collection system. Specific attention should be given to motivate participation among all persons/households in the sample. The field-work staff can be hired and trained, first invitations can be launched and first appointments to the interviews and examinations can be scheduled.

- Stage 5 includes fieldwork and data collection. Some chenges and adaptations to original plans may still be needed, e.g. if participation rates in the first weeks are low or if other problems are faced.

- Stage 6 includes finalizing the data sets, documenting data characteristics and quality, finalizing plans for the data analysis, as well as reporting and disseminating results. One should acknowledge that in the future there will be many analyses not foreseen.

- Quality assurance is essential throughout the survey process (see Part A, Chapter 11 of the EHES Manual).

1.2 Aims and purpose of the survey

Clearly defined and specified aims and purposes guide the survey planning and fieldwork. Time spent in the development of specific aims is time saved in the design of survey instruments and measurements (Biemer & Lyber 2003). There are typically interests to include several topics, instruments and measurements in the survey, but all of them are not feasible due to limited time and other resources, and burden on survey participants. However, the interests of funding agencies need to be considered. The purpose of the survey depends on national needs and uses of health information, but also international action plans, policies and data needs should be considered e.g. collection of data for the European Core Health Indicators (http://ec.europa.eu/health/indicators/echi/list/index_en.htm) and implementation of the WHO Action Plan of the European Strategy (WHO 2012) and the WHO Global Action Plan (WHO 2013) for the Prevention and Control of Noncommunicable Diseases.

Relevant and valid health information is needed for evidence based public policies across several sectors which may influence health, taking into account the Health in All Policies approach (Leppo et al 2013). Information is needed for rational planning and evaluation of health promotion and disease prevention programmes, and health services. In each country the objectives of the survey should take into account ongoing or planned national health promotion programmes and key challenges in developing health services to meet the needs of all population groups. Monitoring and forecasting the population's health and health determinants are prerequisites for sound evidence based public health policy, directing and designing health programmes and services as well as social security. HESs can enhance knowledge on health determinants, health needs and population health, as well as disparities between different population groups. The information from a HES is typical used to:

- assess the prevalence of major diseases and their risk factors as well as related lifestyles and quality of life;
- assess health status and its association with health promotion and disease prevention;
- measure change at an individual (if follow up of the participants is possible) and population level (with regularly repeated surveys);
- predict future health status in the population, based on objective information on major chronic disease risk factors and lifestyles (such as salt and sugar consumption, smoking habits, blood pressure, blood lipid levels, obesity);
- analyse equity in health, health care and wellbeing by providing objective data, comparable in all groups of the population, considering differences by sex, age (especially elderly people), and migrant status or ethnicity;
- estimate met and unmet need for health care, social security benefits and rehabilitation, and to forecast future scenarios concerning the need for health care and social security benefits;
- develop national standards and reference values for the measurements;
- develop a valuable data source for epidemiological studies and health sciences research.

The aims of HESs should be specified and evaluated against other potential sources of health information in each country, such as health interview surveys and administrative and population based registries. This evaluation will show the added potential of HESs to retrieve health information. A HES provides exclusive data on many topics such as disease risk factors not available by any other source (e.g. proportion of asymptomatic conditions). Also, HESs can result in comparable data for many health indicators which are known to differ between countries and between socioeconomic groups. Such differences may relate for example to cultural and social norms and reporting bias. The standardized measurements of health examinations can overcome reporting bias, e.g. the tendency to over-report height and under report weight (Gillium & Sempos 2005, Elgar & Stewart 2008, Tolonen et al 2014). HESs can also reveal shortcomings in the awareness of risk factors, e.g. having high blood pressure (Kastarainen et al 2009, Ostchega et al 2008, Tolonen et al 2014).

HESs provide population prevalence data also in situations where such data cannot be obtained from routine registers because of limited access and use of health services. For example, routine registers reveal diabetes or cardiovascular disease only in those who have used services and been diagnosed (Gnavi et al 2008, Elo & Karlberg 2009). Availability and use of data from administrative databases, electronic health records and disease registers is increasing, but the comparability and usefulness of such data in health monitoring purposes often suffers e.g. from differences in clinical practices, coding systems, access to and use of services and insurance coverage in different population groups and countries (Gavrielov-Yusim & Friger 2014, Thygesen & Ersboll 2014).

When planning and preparing national surveys, possibilities and restrictions in linking register data and administrative databases with HES data should be considered, especially to evaluate non-response and to allow possible follow-up of participants. The use of register data and administrative databases needs to be addressed in the informed consent process (see Part A, Chapter 4 of the EHES Manual).

The scope of the core EHES is limited to the health of the adult working aged population (see Part A, Chapter 2 of the EHES Manual), as both children and the elderly have their own specific health problems, health risks and protective factors, often requiring specific measurements. Surveys among children and the elderly also have their own challenges in regard to survey ethics and fieldwork practices, which is why the EHES standards are at first targeted to adult health surveys. The EHES survey can be extended to also cover the elderly as the core measurements are feasible with similar methods among the elderly, but their specific needs should be taken into account (e.g. inclusion of institutionalized persons, scheduling appointments, and consent among those with cognitive disabilities).

The scope of the core EHES is limited to the health of the adult population, as both children and the elderly have their own specific health problems, health risks and protective factors, often requiring specific measurements. Surveys among children and the elderly also have their own challenges in regard to survey ethics and fieldwork practices, which is why the EHES standards are at first targeted to adult health surveys. The EHES survey can be extended to also cover the elderly as the core measurements are feasible with similar methods among the elderly, but their specific needs should be taken into account (e.g. inclusion of insitutionalized persons, scheduling appointments, and consent among those with cognitive disabilities). Age-group specific measurements and other additions will be developed later and included in the EHES Manual.

1.3 Implementing EHES standards in national surveys

Countries have three alternatives for implementing the EHES standards: (Tolonen et al 2008):

1. **Building a new national HES.**

   When a new national HES without any (or recent) prior HESs in the country is organized, planning and implementation of the survey should be based on the EHES standards. Needs to adapt the European standards to the national circumstances should be considered. National experts need to decide which of the options in this manual are most feasible in their country, taking into account how these choices affect the comparability of data.

2. **Synchronizing EHES standards with the existing HES.**

When synchronizing EHES standards with an existing national HES it may be challenging to balance between national time trends and European comparability. A specific pilot study may be needed to compare results from examinations carried out by different protocols. Some measurements and/or questions may need to be administered to the same respondents in two different ways.

3. **Combining national HIS and EHES.**

   When combining the EHES with a national HIS (EHIS), the challenge is in organizing the data collection successfully and minimizing selection bias if the HES part is organized separately after the interviews.. Everybody in the sample (i.e. not only the respondents of the HIS) should be invited to the HES whenever possible (e.g. when the EHIS sample is not too large to be feasible for the HES) . There are several examples showing that inviting only the participants of previous phases leads to a diminishing participation rate for HES. The HES and HIS can be used for complementing missing HIS information in HIS non-participants, and the HIS can be used to evaluate non-response in HES. If the HIS is collected by a different organization (e.g. National Institute of Statistics) than the HES, these surveys need to be prepared in close collaboration. In addition to differences in sample size and sample selection, many practical details in the data collection need to be considered.

A HES always includes one or several self-administered questionnaires and/or interviews. These may be very extensive and time-consuming, e.g. when including the full EHIS questionnaire (EHIS 2013). It is important to consider the respondent burden related to interviews and self-administered questionnaires, and their effect on the persons' willingness to participate.

Some countries may collect HES information through national health screening or primary health care services by inviting persons to examinations carried out in primary health care facilities. Key issues in the feasibility of utilizing regular screenings and primary health care organizations for national health monitoring purposes are their coverage at population level (assuring the representativeness and avoiding selection bias), and standardization of the measurements (e.g. local premises, equipment and adequate training of personnel), as well as the other components of quality assurance (see Part A, Chapter 11 of the EHES Manual).

A modular structure can be considered if the survey covers seveal additional topics which are not relevant or feasible to all population groups (Figure 1.2). These modules will need to be taken into account in the survey management and fieldwork logistics. There may be additional measurement modules e.g. on functional ability for those aged 65 and over. An additional measurement module for a sub-sample may include e.g. a time consuming interview or examination which are not feasible for the total sample due to limited resources or respondent burden (e.g. the FINDIET survey in the FINRISK studies, Reinivuo et al 2010,

and the mental health module in the German national HES, Jacobi et al 2013).

Total sample

EHES Core measurement

Sub-sample of the total sample

Additional measurement module A

Sub-sample of the total sample

Additional measurement module B
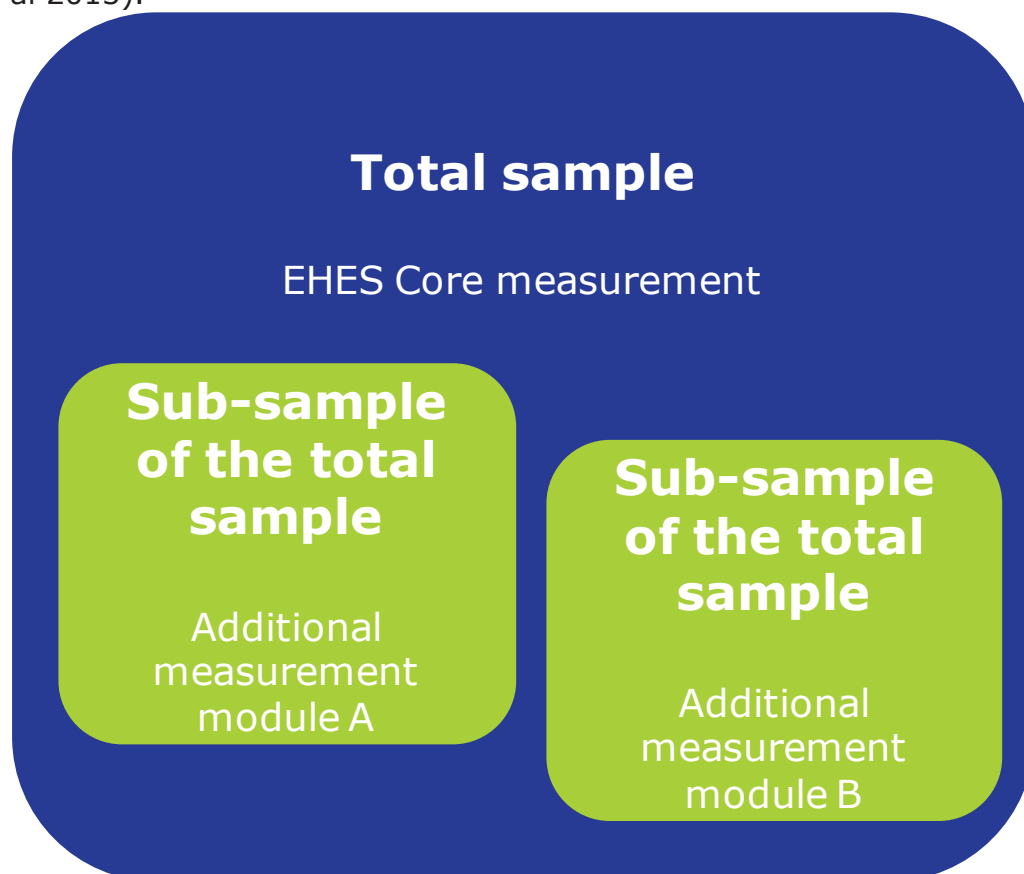
Figure 1.2 Example of a modular structure in the survey

1.4 Survey management

An interdisciplinary survey team is needed for the planning, design, implementation and evaluation of the survey. A core group of key experts should ensure that different aspects are taken into account. In addition, many other experts are needed, and within larger survey organizations their work needs to be organized in different teams, led by members of the core group or others closely involved in the survey. In smaller survey organizations various experts may be consulted without involving them in the actual survey organization. Various types of expertise should be utilised:

- Policy experts to define the needs and use of data for evidence based policy making and to use the results for these purposes;
- Health care and other public service professionals to define the needs of data for planning and evaluating health services and health promotion activities and to use the results for these purposes;
- Scientists in the fields of epidemiology, statistics, public health, and other health and social sciences to define the use of the data for scientific research purposes;

- Expertise in fieldwork logistics and supervision, laboratory issues, data management, information technology, communication and dissemination is needed to make sure that the data collection runs without problems and to ensure high quality data.

It is also useful to involve different stakeholders such as ministries (e.g. health and research), social insurance organizations, and non-governmental organizations to express their interests for the survey, to promote the survey for fund raising and raising interest among the population to participate, and to disseminate the results.

1.4.1 Management structure

The organizational responsibilities of a HES can be divided into four partly overlapping key tasks (adapted from Tolonen et al 2002):

1. **Planning:** Definition of the objectives and scope of the survey, planning and preparing the fieldwork and other survey operation.
2. **Operation:** Implementation and operation of systems for data collection (fieldwork) and data processing.
3. **Quality assurance:** Monitoring performance, providing feedback, and ensuring that the results are within predefined quality limits.
4. **Dissemination:** Making sure that information about the survey is widely available for the target population as well as for the stakeholders. This may be crucial to ensure adequate participation, and utilization of survey data and the results.

Planning and operation are most often lead by the same organization, while in some countries e.g. the Ministry or National Public Health Institute are responsible for planning while the organization responsible for the operation is selected from competing organizations such as universities and other research organizations. This may lead to difficulties in standardization and in ensuring that the experiences and expertise of key personnel can be kept up. Close collaboration with local organizations (e.g. primary health care centres) and personnel (GPs, local authorities) is of paramount importance in each phase: for minimizing the costs (e.g. offering clinic space), for motivating the population and improving the participation rate, and for giving advice based on the participant's personal measurement and blood test results. It also needs to be decided if there is a need to carry out the quality assurance by an organization or persons without vested interest in the national survey, but with adequate knowledge of the process and methods.

A clear management structure of the survey helps to:

- ensure that the set objectives can be met;

- make planning and implementation of the survey more efficient;
- increase the quality of the entire survey;
- decrease the cost of the entire survey.

An example of the management structure of a national HES is given in Figure 1.4. The tasks of different groups and persons will depend on the national health care system. In the example, the different groups and persons have the following tasks:

A. The Steering Committee (or a Steering Group) approves the survey objectives, and provides directions and guidelines to meet these aims.

B. The Project Manager runs the survey. He/she is responsible for:
- the organization of the survey by allocating responsibilities and resources and by making sure that all areas are covered and that there is no overlap between the responsibilities of different experts;
- managing the survey process by making decisions, giving guidance, providing and acquiring assistance, motivating team members and solving possible conflicts;
- day-to-day monitoring and evaluation of the survey process, schedules and budget and making adjustments to these when needed;
- reporting to the Seering Committee.

C. The Core group assists the Project Manager. It consists of key experts, selected from the Team Leaders or other experts, with specific responsibility for coordination of fieldwork, statistical issues, and data management.

D. Survey Teams: Different subareas of the survey are planned and implemented in larger surveys by different Survey Teams, led by the Team Leaders. In smaller surveys there may be only single experts in each area, or one expert is covering several areas of expertise. These teams or experts cover different areas of expertise, such as sampling, fieldwork, laboratory issues, communication and quality assurance, as well as different topics of the survey (e.g. blood pressure monitoring, obesity). The opic specific teams or experts can evaluate and propose questionnaire instruments and measurements for their areas of expertise, as well as plan and carry out special studies (modules). When fruitful collaboration is built during the planning and preparation, the members of these expert groups are a valuable resource for e.g. training of the fieldwork staff, quality control during the fieldwork, data analysis and reporting, as well as ensuring wide use of the survey data for several research purposes.

```
┌─────────────────────┐
│   Project Manager   │
└──────────┬──────────┘
           │        ┌─────────────────────┐
           │────────│   Steering Group    │
           │        └─────────────────────┘
┌──────────┴──────────┐
│  Core Group/ Survey │
│        Team         │
│                     │
│ • Fieldwork coordinator
│ • Data manager      │
│ • Statistician      │
│ • Laboratory specialist
│ • Communications officer
│ • Other experts     │
└─────────────────────┘
```
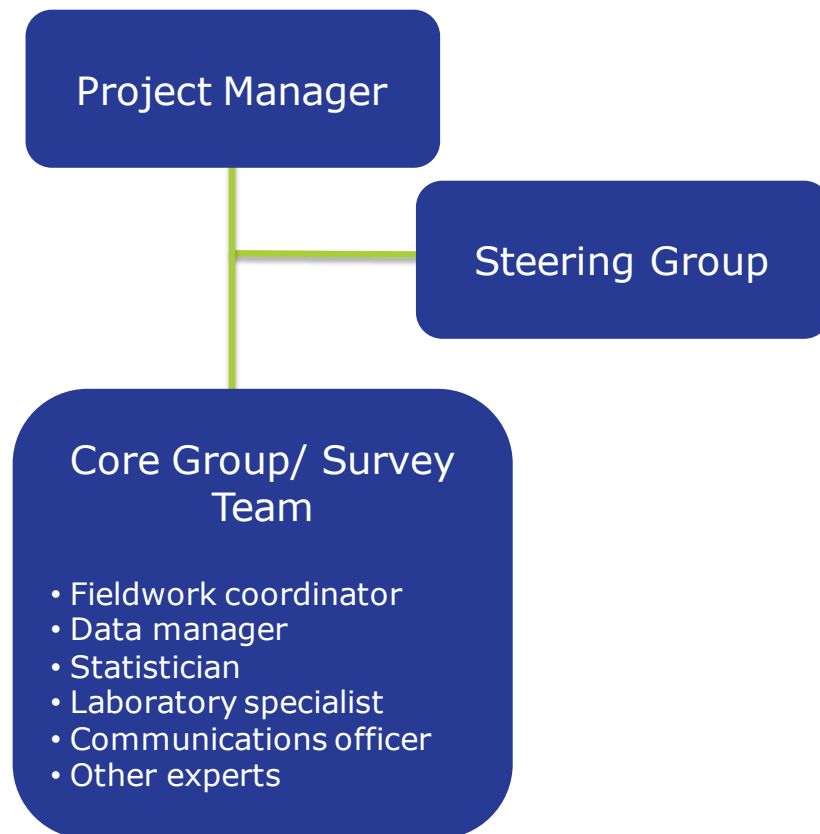
Figure 1.3 An example of a survey project organization in a survey including EHES core measurements

Key experts and tasks in the survey project organization include:

- a fieldwork coordinator or fieldwork team responsible for the fieldwork logistics, training and day to day data collection activities. In larger studies a full time fieldwork coordinator is needed to share the workload of the Project Manager;

- a data management expert, when needed supported by the IT team. They are responsible for the computer systems and programs, and the data management;

- a person responsible for the laboratory activities, when needed supported by the laboratory team responsible for the sample collection, analysis and storage;

- in larger studies a quality assurance team may be needed for the quality assurance activities;

- a survey statistician or a team of statisticians with specific expertise on sampling or survey data analysis;

- a person with expertise in survey ethics may need to be consulted or invited to the fieldwork team;

- a communications specialist may be needed e.g. to update the website and to utilize social media during the fieldwork and to spread information about the survey results.

Some of the tasks may be carried out under a short-term contract (e.g. computer systems, data entry, printing, mailing) or by contracting out some functions to an external organization. The roles and responsibili-

ties of these persons/teams may vary between countries due to legislation and differences in organization structures. Legislation in many countries calls for a chief physician in any study classified as medical research. The roles of the fieldwork co-ordinator and chief physician mey be combined.

If the survey team is large and if the survey covers different data collection phases, and/or several topics or modules, it may be useful to have special teams devoted also to each topic area (Figure 1.4).
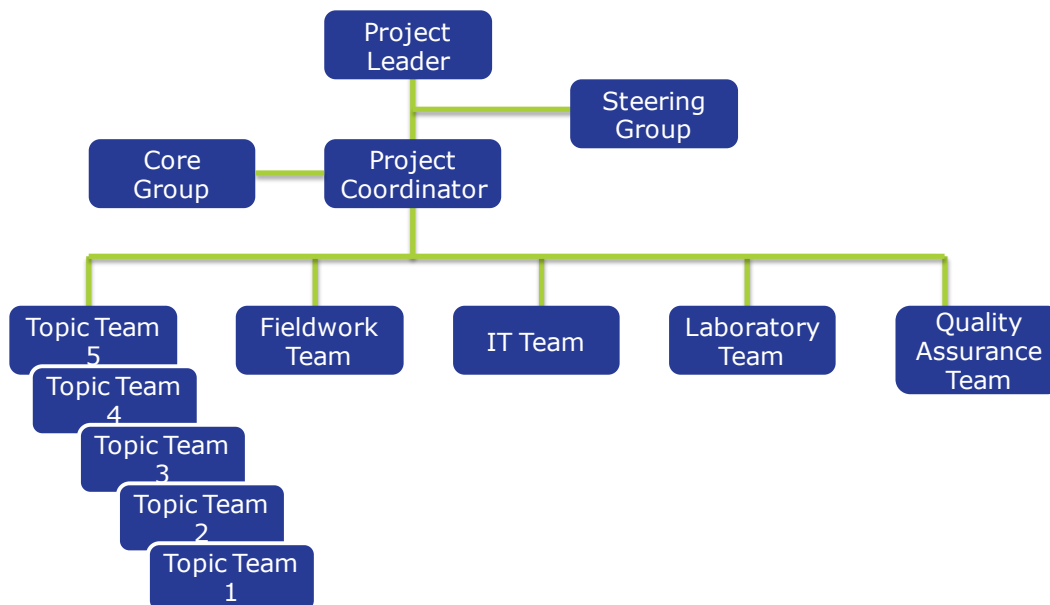
Figure 1.4 An example of a survey project organization in a comprehensive survey including EHES core measurements and several additional measurements

1.4.2 Management tools

It is essential to ensure that there is enough time for different phases of the survey process. The planning and preparation will usually require at least one year before the fieldwork can be started (Figure 1.5). If there is no recent (within last 5-10 years) or only little experience of a previous survey in the country, the planning and preparation for a full scale HES requires a longer period of time.

As collaboration between several organizations, teams and experts is needed, the detailed planning and preparation may benefit from using specific project management tools and software to define the project timeline and to follow progress.
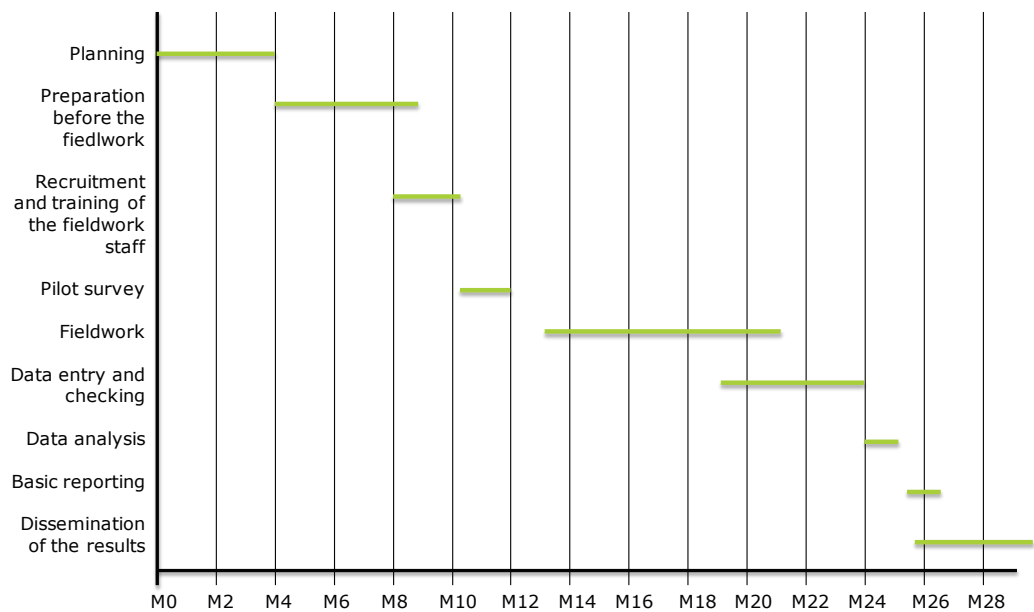
Figure 1.5 Example of the timeframe for the survey

One key element in the survey process, to ensure a successful data collection and fieldwork phase, is piloting and detailed evaluation of the pilot process. Sometimes a small pre-pilot (e.g. fieldwork testing with volunteer participants) is needed before the full pilot to test the computer programs, measurement techniques and timing. Specific aims for the pilots need to be defined during the planning and preparation. A pilot phase is always recommended, but the aims and content of the pilots depend on previous experience and frequency of the survey. Adequate time between the pilot and the actual data collection should be ensured so that the experiences and results of the pilot are evaluated in detail and the needed specifications and modifications are made to the programs and manuals (see Part A, Chapter 11 of the EHES Manual).

1.4.3 Risk analysis

The aim of risk analysis is to avoid uncertainties that threaten the goals, timetables and budget of the project, and to take actions in advance to reduce the effect of these risks. Risk analysis should be carried out when planning the project and updated during the process. An example of risk analysis is presented in Table 1.1. Some of the risks may be rather minimal and easy to solve, e.g. the risk for crisis situations is very minimal and the loss of data can be prevented by proper data management. The very minor risks or discomfort caused to participants during blood sample drawing can be prevented by choice of competent personnel and their training before fieldwork. Risks related to adequate time and personnel resources can be more challenging, if adequate funding is not available. Several other rare risks can be anticipated, e.g. violation of personal data protection rules, which can be prevented by data management, quality control and training of personnel.

Table 1.1 Risk analysis in a national HES, examples of potential risks

Risk	Problems caused	Options for avoiding and controlling the risk
Insufficient personnel resources for planning and preparation	Shortcomings in planning and preparation leading to problems during fieldwork, in standardization and quality of data	Careful preparation of the survey organization, and seeking mandate from the ministries (health and research). Seeking specific funding for the planning and preparation, careful budgeting and diverse fund raising (see Part A, Chapter 16 of the EHES Manual) to ensure that the needed resources are available.
Shortage of fieldwork personnel	Difficulties in keeping time schedules: problems caused for participants as well as in getting results	Careful piloting and planning for the time schedules, taking potential sick leaves and turnover of personnel into account when planning the size of fieldwork team(s) and in the training programmes.
Insufficient time between pilot and actual fieldwork	Not possible to correct errors, specify manuals and training or adapt protocols, leading to problems in standardization and data quality	Acknowledging the aims and significance of the pilots.
Problems in collaboration between different organizations and actors	Difficulties in utilizing all expertise needed, and problems in keeping time schedules	Well defined leadership, building partnerships throughout the survey process, careful planning for the supervision of the fieldwork teams
Low motivation among the population to participate	Low response, selective participation, biased results	Media campaigns and careful planning of the recruitment process (see Part A, Chapters 13 and 14 of the EHES Manual)
National or local political or ecological crisis situations	Loss of data	Timely data transfer to central data centers.

Risk	Problems caused	Options for avoiding and controlling the risk
Epidemics	Absences of fieldwork staff, difficulties in participation	Little possibilities to avoid: infectious disease control at fieldwork settings and offering seasonal flu vaccinations to fieldwork staff.
Safety risks during fieldwork	Harm caused to staff members or participants	The protocol for needle stick injuries should be easily available to all staff members at all examination sites. Safety risks during fieldwork covered in manuals and training. Adequate supervision of field work staff throughout the fieldwork process.

1.4.4 Project evaluation

Project evaluation should be an ongoing task (Table 1.2). It helps to make sure that the survey will be finalised with the resources available and within the timeframe set for the survey. Some parts of the evaluation are directly linked with quality assurance. Indiicators for evaluation should be defined and followed with regular intervals and actions developed if the targets (e.g. numbers of participants) are not met.

Table 1.2 An example of potential evaluation indicators for selected stages in the survey process

Survey stage	Process indicators program operations	Output indicators direct results or products of project activities	Outcomes indicators impacts or changes that can be attributed to the project activities
Survey design	Organized meetings and seminars	First version of the survey proposal	National consensus on carrying out the HES and timing of the surveys. HES plans approved by national authorities with at least preliminary decisions for funding for the HES.
Planning and preparation	Number and type of experts involved in the survey planning, personnel resources needed	Detailed survey plan with a budget	Ethical approval
Fieldwork during pilot(s) and the actual survey	Training seminars organised for the fieldworkers: hours of training Number of invited persons	Number of fieldwork staff members who participated in the full national training (% of all fieldworkers) Number of days for the fieldwork Numbers of participants, those who were found to be ineligible, those who were not contacted and those who refused (by age and gender) Recorded length of examinations per participant – reported average length per participant (minutes/ hours) Place of examinations: number of participants examined (if needed specify at the clinic setting/at home/ at an institution)	Participation rate (per age/gender) Cost of the survey data collection/participant

References

- Biemer PP, Lyberg LE. Introduction to Survey Quality. Wiley Series in Survey Methodology, John Wiley & Sons, New Jersey, USA, 2003.
- Czaja R, Blair J. Designing surveys. A Guide to Decisions and Procedures. 2nd edition, Sage Publications, Inc., California, USA, 2005.
- European Health Interview Survey (EHIS wave 2) - Methodological manual - 2013 edition. http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-13-018

- Elo SL, Karlberg IH. Validity and utilization of epidemiological data: a study of ischaemic heart disease and coronary risk factors in a local population. Public Health. 2009 Jan;123(1):52-7.

- Elgar FJ, Stewart JM. Validity of self-report screening for overweight and obesity. Evidence from the Canadian Community Health Survey. Can J Public Health. 2008 Sep-Oct;99(5):423-7.

- Font J, Mendez M. Surveying Ethnic Minorities and Immigrant Populations. Amsterdam University Press, Amsterdam 2013.

- Franklin S, Walker C (eds). Survey methods and practices. Statistics Canada. Social Survey Methods Division, 2003. Also available at http://www.statcan.gc.ca/pub/12-587-x/12-587-x2003001-eng.pdf

- Cavrielov Yusim N, Friger M. Use of administrative medical databases in population-based research. J Epidemiol Community Health 2014;68:283-287.

- Gillum RF, Sempos CT Ethnic variation in validity of classification of overweight and obesity using self-reported weight and height in American women and men: the Third National Health and Nutrition Examination Survey.Nutr J. 2005 Oct 6;4:27.

- Gnavi R, Karaghiosoff L, Costa G, Merletti F, Bruno G. Socio-economic differences in the prevalence of diabetes in Italy: The population-based Turin study. Nutrition, Metabolism and Cardiovascular Diseases, 2008, Volume 18, Issue 10, Pages 678-682.

- Jacobi F, Mack S, Gerschler A, Scholl L, Höfler M, Siegert J, Bürkner A, Preiss S, Spitzer K, Busch M, Hapke U, Gaebel W, Maier W, Wagner M, Zielasek J, Wittchen HU. The design and methods of the mental health module in the German Health Interview and Examination Survey for Adults (DEGS1-MH). Int J Methods Psychiatr Res. 2013 Jun;22(2):83-99. doi: 10.1002/mpr.1387.

- Kastarinen M, Antikainen R, Peltonen M, Laatikainen T, Barengo NC, Jula A, Salomaa V, Jousilahti P, Nissinen A, Vartiainen E, Tuomilehto J. Prevalence, awareness and treatment of hypertension in Finland during 1982-2007. J Hypertens. 2009 Apr 30.

- Leppo K, Ollila E, Peña S, Wismar M, Cook S. (Eds) Health in All Policies. Seizing opportunities, implementing policies. Ministry of Social Affairs and Health, Finland, 2013. (http://www.euro.who.int/__data/assets/pdf_file/0007/188809/Health-in-All-Policies-final.pdf)

- Ostchega Y, Hughes JP, Wright JD, McDowell MA, Louis T. Are demographic characteristics, health care access and utilization, and comorbid conditions associated with hypertension among US adults? Am J Hypertens. 2008 Feb;21(2):159-65.

- Primatesta P, Allender S, Ciccarelli P, Doring A, Graff-Iversen S, Holub J, et al. Cardiovascular surveys: manual of operations. Eur J Cardiovasc Prev Rehabil. 2007 Dec;14 Suppl 3:S43-61.

- Rechel B, Mladovsky P & Deville W. Monitoring migrant health in Europe: a narrative review of data collection practices, Health policy 2012; 105 (1), 10-16.

- Reinivuo H, Hirvonen T, Ovaskainen ML, Korhonen T, Valsta LM. Dietary survey methodology of FINDIET 2007 with a risk assessment perspective. Public Health Nutr. 2010 Jun;13(6A):915-9. doi: 10.1017/S1368980010001096.

- Thygesen CL, Ersboll AK. When the entire population is the sample: strengths and limitations in register-based epidemiology. Eur J Epidemiol 2014; 29: 551-558.

- Tolonen H, Kuulasmaa K, Laatikainen T, Wolf H, for the European Health Risk Monitoring Project. Recommendations for indicators, international

collaboration, protocol and manual of operations for chronic disease risk factor surveys. 2002 Available from: http://www.thl.fi/publications/ehrm/product2/title.htm

- Tolonen H. Towards the High Quality of Population Health Surveys. Standardization and Quality Control. National Public Health Instiute A27/2005, Helsinki 2005; Available from: http://urn.fi/URN:ISBN:951-740-546-4

- Tolonen H, Koponen P, Aromaa A, et al. (Eds.) Recommendations for the Health Examination Surveys in Europe. B21/2008, Publications of the National Public Health Institute, Helsinki 2008. Also available from http://urn.fi/URN:ISBN:978-951-740-838-7

- Tolonen H, Koponen P, Mindell JS, Männistö S, Giampaoli S, Dias CM, Tuovinen T, Gößwald A, Kuulasmaa K; European Health Examination Survey Pilot Project. Under-estimation of obesity, hypertension and high cholesterol by self-reported data: comparison of self-reported information and objective measures from health examination surveys. Eur J Public Health. 2014 Dec;24(6):941-8. doi: 10.1093/eurpub/cku074. Epub 2014 Jun 6.

- WHO. Action Plan for implementation of the European Strategy for the Prevention and Control of Noncommunicable Diseases. World Health Organization, Regional Office for Europe, 2012. Available from http://www.euro.who.int/en/publications/abstracts/action-plan-for-implementation-of-the-european-strategy-for-the-prevention-and-control-of-noncommunicable-diseases-20122016

- WHO. Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013-2020. World Health Organization, 2013. Available from http://www.who.int/nmh/events/ncd_action_plan/en/
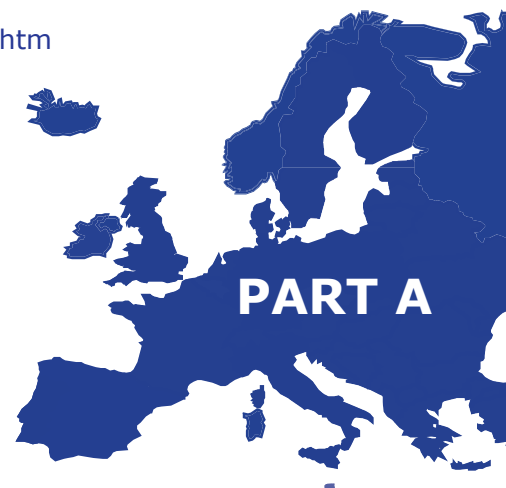
2. Target population and sample size

PART A

Johan Heldal[1], Susie Jentoft[1], Kari Kuulasmaa[2]

[1] Statistics Norway (SSB), Oslo, Norway
[2] National Institute for Health and Welfare (THL), Helsinki, Finland

2.1 Target population and sample size

The target population is the population of individuals which we are interested in describing and making statistical inferences about. EHES Refernence Centre (RC) suggests the following definition for the target population of a country:

1. The core target population is the set of all persons aged at least 25 years and at most 64 years and having permanent residence in the country. (Instructions for more precise definition of age are given in Part A, Chapter 3.)

2. Each country can extend the eligible age group with a lower bound of 18 years and with no limitation for the upper bound. The core measurements are very relevant also for those below 25 or above 65 years of age. Those below 25 were not included in the core target population because of a very practical reason: In most countries, it is very difficult to achieve reasonably high response rates of the young adults, and therefore their survey results can be unreliable. For the elderly there may be different tests with a higher priority. In the eldest age groups (75 and over) some ethical and practical issues may be more complicated because of a relatively high percentage of institutionalized persons in some countries and persons with limited cognitive functioning. Therefore, the elderly age group is not included in the core adult survey defined in this manual.

Some counties have already defined the age ranges for their surveys wider than this. However, EHES is currently a survey of adults. The importance of children and adolescents as a target group is acknowledged but the current recommendations are limited to adults. Inclusion of children and adolescents to the study would require further attention to ethical issues. Furthermore, children and adolescents have other

relevant health issues, for example relating to growth and development, which require specific measurement procedures.

The above definition of the target population describes the ideal target. However, it is recognized that it is not always feasible to include all population subgroups in the national HES. In order to take a sample from a population one needs a sampling frame from which a sample can be taken. Some countries will have difficulties establishing sampling frames that cover the entire population at a specific date. For example, in some of the health examination surveys carried out so far, institutionalized persons have not been available for sampling. There are different types of institutions. Among the most common are nursing homes, elderly homes, military barracks, jails and monasteries. Also students living in halls of residence may not be included in the sampling frame. Each of these should be considered separately taking into account the facts that they may require special designed sampling frames. Furthermore, it is sometimes difficult to contact people living in institutions even if they are in the sampling frame. In some countries, non-citizens may not be available in the same sampling frame as the citizens even if they live permanently in the country.

It is not recommended to exclude form the survey population groups difficult to contact or who do not for example speak major languages of the country. Nevertheless, it is recognized that such population groups may require specific protocols and extra resources, such as home visits and interpreters which need to be considered when defining the national target population.

Every detail of the national target population and the coverage of each national sampling frame must be well documented in order that the results can be interpreted correctly.

Selection of sampling frames is discussed in Part A, Section 3.2.

2.2 Sample size

A minimum of 4000 persons are sampled to be invited in each country. Each of eight age-sex domains (25-34, 35-44, 45-54 and 55-64 years) should have at least 500 representatives in the sample. The sample size calculation is based on a participation rate of 70 percent, but should be applicable also if the realistic expectation for the participation rate is different. This minimum size relates to the requirements for statistical power when testing differences between countries for age-gender domains. For comparisons between regions or socioeconomic groups, or by ethnicity (or country of origin) , each country will have to set its own standards for accuracy and explain its needs for larger sample sizes.

Sample size relates to the statistical precision of the survey results, whereas bias is a concern related to low response rate. The relative benefit from higher precision, and therefore higher number of participants, is better if the response rate is high. On the other hand, if the expected response rate is low, it will be better to spend resources on increasing the response rate than to increase the total sample size. Specifications and calculations of the minimum recommended sample

size are given in the FEHES recommendations (Tolonen 2008). In the pilot survey we recommend a sample size of at least 200 persons, with at least 25 persons in each age-gender domain.

If the elderly are included in the survey, attention should be paid to ensure that the sample size will be sufficient in the oldest age groups even though the population size is smaller than in the younger age groups. This may require oversampling, i.e. increased sampling probabilities in the oldest age groups. The same can be considered also for other relatively small population groups of specific interest, such as the migrants.

How to obtain a sample is discussed in Part A, Chapter 3.

2.3 The EHIS definition

For the European Health Interview Survey (EHIS) the Task Force III report on sampling issues (Axelson 2009) suggests the following definition of the target population. (Note, however, that a new regulation on EHIS is in preparation and may bring an update to the EHIS sampling recommendation.)

For the EHIS, the target population should contain all adults (15 years old and over) living in the country at the place of their usual residence (the place where they mainly live). The sample may not include individuals at a place of residence where they do not mainly live. Such individuals must be treated as not eligible, and the interview stated as terminated.

Apart from defining a wider age range, this definition does not specify a reference point in time for the target population. We have not seen a reference to 'place of their usual residence' as relevant. The EHIS definition may seem to exclude persons not having a residence (homeless people). It is not clear whether this is intentional. However, such persons can be difficult to reach and are not likely to be interviewed or examined anyway.

References

- Tolonen H, Koponen P, Aromaa A, Conti S et al (eds.) Recommendations for the Health Examination Surveys in Europe. Publications of the National Health Institute, Helsinki 2008. B21/2008. Available athttp://urn.fi/URN:ISBN:978-951-740-838-7. Accessed on 22 January 2013
- Axelson M, Bihler W, Djerf K, et al (prepared by) European Health Interview Survey. The Task Force III report on sampling issues. 2009. Available at https://circabc.europa.eu/sd/d/4f89b382-b3e1-4884-94e1-ca9f2e2338b6/Task%20Force%20III%20report%20on%20sampling%20issues%20up%20to%20date%20November%202009.pdf.

3. Sampling procedures

PART A

Johan Heldal[1], Susie Jentoft[1]

[1] Statistics Norway (SSB), Oslo, Norway

The goal of a Health Examination Survey (HES) is to produce statistics for clinically measured heath indicators, such as the average state and variation in various health indicators for national populations. This should be done in such a way that the estimates obtained from the survey are as well as possible statistically unbiased for the true averages of these indicators in each participating country. This is required if we want to be able to compare estimates among countries and carry out unbiased tests for these differences. There are many potential sources of bias in a health examination survey. Sampling bias is one of them, but it is also one of the sources of error that can be brought almost entirely within control of the survey taker. At the same time as avoiding bias, purely random errors in the estimates should be made as small as possible.

Control of sampling errors, both systematic (bias) and random, requires a good *sampling design*. Scientific surveys make use of *probability sampling*. This means that every eligible individual or household should have a known probability of being sampled. In probability sampling, randomization techniques and (pseudo) random mechanisms are used to select the individuals to be invited to the survey. Survey sampling is a science which should be carried out or monitored by professional statisticians in each country. The procedures for estimating the health indicators rely on probability sampling.

There are many ways to select a probability sample. Which method to choose will depend on the features of the actual survey and the sampling frames that are available (see Part A, Section 3.2 of the EHES Manual). A two stage sampling design is recommended for health examination surveys in all countries except possibly the smallest ones. Depending on the sampling frame available, more than two stages might be necessary in some countries in order to reach down to the individual people to be invited to participate in the survey.

The EHES RC has developed a sampling application program to simplify sampling in line with the recommendations given in this chapter.

The program is written as an R-package called EHESsampling (Jentoft 2011). R is freeware and can be downloaded from http://www.r-project.org/. EHESsampling and the user manual can be downloaded from the EHES web site (http://www.ehes.info). EHESsampling carries out the steps described in Part A, Sections 3.3.4 - 3.3.6 of the EHES Manual and contains the recommendations in these sections as default options.

Part A, Section 3.1 of the EHES Manual is a short overview of some general considerations concerning Health Examination Surveys and their implications for design of such surveys compared to Health Interview Surveys.

Part A, Section 3.2 of the EHES Manual discusses some alternative sampling frames for a HES. Section 3.3 treats the design for Stage 1 (sampling PSUs) in further detail and how efficient sample sizes can be calculated. Section 3.4 discusses the design for Stage 2. Section 3.5 considers aspects of using address frames. Section 3.6 deals with documentation and data management and Section 3.7 discusses procedures that are common in use but not recommended here.

3.1 General considerations

3.1.1 Health Examination Surveys

In a two-stage sampling design the sample of individuals to be invited is obtained after two stages of sampling. In a Health Examination Survey where participants are invited to an examination site, it is essential that the distance to the clinic is as short as possible. Short distances are also important when mobile clinics are used to visit invitees closer to their homes (e.g. in rural districts) or when there are home visits by field work staff (See Part A, Chapter 7 of the EHES Manual for a more detailed discussion). Both situations call for a clustering of the invitees in a limited number of examination areas that do not cover an entire country but may be selected from a larger set of potential examination areas which do cover the entire country. In agreement with the general terminology of survey sampling, the examination areas will be called *Primary Sampling Units* (PSUs) in this chapter. The design for sampling of PSUs is Stage 1 of the total sampling design. The individual participants are sampled from the PSUs that have been selected at Stage 1 (see Figure 3.1). Within each PSU being sampled at Stage 1, at Stage 2 we will sample people, addresses, households or dwellings. These units will be termed *Secondary Sampling Units* (SSUs).

EHESsampling application produces R data frames with information on the sampling frame as well as for the sample itself for each stage of sampling in all strata, such as stratum and PSU identification and inclusion probabilities for sampling units at each stage. It is essential that all this information is stored both for those who participate and those who do not along with the data collected in the survey. No information should be discarded. This information is needed for proper analysis, estimation and variance calculations later. All details of the national

sampling designs and samples resulting from them must be well documented.
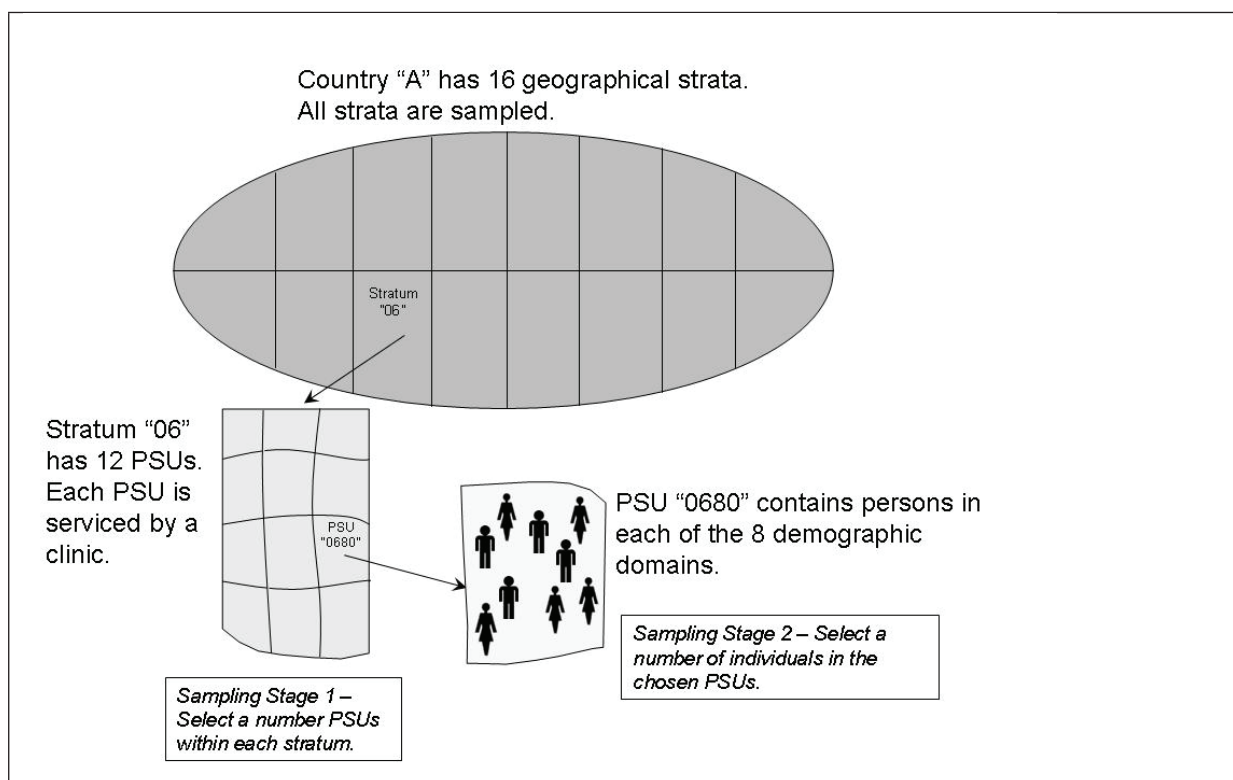
Figure 3.1 The principle for a two-stage design

3.1.2 The recommendations for European Health Interview Survey (EHIS)

For those involved also in EHIS, we describe here its sampling recommendations and the differences from the EHES sampling.

The recommendations for sampling in EHIS are given in the EHIS Task Force III report on sampling issues (Axelson 2009). (Note, however, that a new regulation on EHIS is in preparation and may bring an update to the EHIS sampling recommendations.)Since EHIS is an interview survey only, its sampling design does not need to pay attention to such things as "closeness to a clinic" for the participants. It can even be carried out by telephone or as self-administered survey in which case a two-stage design does not have any advantage. The way HIS surveys have been carried out, as well as survey design, differs between the European countries. The recommendations for sampling in EHIS TF report are therefore not very specific. Basically they recommend that samples should be taken as probability samples where each member of the target population is assigned a non-zero probability of selection.

Some countries have expressed interest in coordinating HES with the HIS, for example conducting HES on a subsample of HIS. Such a strategy requires a sampling design for HIS which is compatible with the requirements for HES. This is not the case for all of the HIS surveys that have been carried out so far. However, it may be possible to coordinate the two surveys in the future.

The EHIS TF III report discusses pros and cons with substitution of non-responding sample units and some practices that occur in the EU Member States. We will not repeat all the details of the discussion here, but in conclusion, it recommends not using substitution of non-respondents in population health surveys. In EHES as well as EHIS we must expect that the state of health of the invited people will often be a factor causing non-participation among the invitees. This will bias the estimates from the survey. Substituting respondents will most likely have of similar health as other respondents and will therefore not reduce this bias. On the contrary, under some circumstances it may increase the bias. Moreover, when substitutions are being used the inclusion probabilities, the probabilities of being selected to the sample, can no longer be exactly calculated. For these reasons we recommend *not* to use substitutions in HES. Observations identified as substitutions will be excluded from the final comparative analyses of the EHES data. Therefore, if a country still chooses to use substitutions these must be identifiable in the data. However, reason for non-participation should be recorded in detail.

Readers interested in more details of former HIS surveys can consult the webpage https://hishes.wiv-isp.be/

3.2 Sampling frames

When a survey is carried out in more than one stage, a sampling frame for each stage will be required. The frame for Stage 1 should be a list of all PSUs that can be selected with information on their population sizes (people or households/dwellings). If updated statistics are not available for all PSUs the best available estimates, e.g. the last census, can be used. If feasible, the population sizes should be broken down to at least the core age by sex groups for which statistics will be published in EHES. There should also be information about which stratum each PSU belongs to. The PSUs and the strata should be equipped with a unique digital number as well as names.

The Stage 2 sampling frame is the list of units (individuals or addresses/households/dwellings) from which a sample of such units can be taken. If the list contains individuals, Stage 2 will be the final stage. If the units are addresses there may be a need for a Stage 3 to select dwelling in a multi dwelling house. Stage 3 sampling will often have to take place in the field and will not be covered in detail in this chapter. A Stage 2 frame should be established at least for all the PSUs selected at Stage 1. It is recommended that the Stage 2 frame is updated as closely as possible to the time when the PSU will be visited by the survey. It can therefore be an advantage waiting as long as possible before taking the Stage 2 sample for a selected PSU. Availability of high quality sampling frames for Stage 2 differs among countries. While some countries have central population registers that can be used in other countries sampling frames for Stage 2 are only available at the local level, e.g. municipalities. Different kinds of frames for Stage 2 are recommended in the following order

1. Whenever legally and practically available, a central file with the most recent and best coverage of the *people* in the target population should be used as the sampling frame. Ideally, this will be a population register. If possible, the main frame can be supplemented with other files to catch parts of the target population not covered by the main frame. Some countries have frames covering individuals but lack for instance non-citizens, homeless or parts of the institutionalized population. The extent of such under-coverage by cause should be estimated. See below. Most European countries have a census every ten years. Fresh census data is very useful as a sampling frame and should be considered for the national HES.

2. If a quality frame with individuals is not available, an updated address file or list of housing units can be used as an alternative. However, a postal address can either address a dwelling directly or a house with many dwellings. The two situations require somewhat different approaches to sampling.

3. Countries already carrying out national HES with samples drawn from an established frame may continue to use the same frame in the future. However, all such frames must be compared and evaluated against the general recommendations and standards proposed for EHES.

4. Countries that do not have a sampling frame mentioned in 1 or 2, can construct a Stage 1 sampling frame based on available statistics for the units chosen for Stage 1 and the sampling of such units can be carried out in the same way as for countries covered by item 1 or 2. Some countries have local population registers which can be frames for Stage 2 sampling when the Stage 1 sample has been selected. If these kinds of local frames are not available, a local frame must be constructed. It may be necessary to sample in more than two stages. The strategies may differ between urban and rural areas. In cities, street maps which identify city blocks may be useful. The number of dwellings in each block must be mapped and some of them sampled. Dwellings can then be sampled within each selected block. In rural areas it may be better to use areal squares as PSUs. The number of houses in each square should be counted and a sample of the inhabited squares selected. A sample or all houses in the sampled squares should be included in the sample. This is called an area frame. The National Health and Nutrition Examination Survey in the USA and the Canadian Health Measure Survey use this kind of strategy. See the NHANES Sample Design (Johnson 2014) for descriptions. Each country needing this kind of frame must adapt a procedure that fits the national structures.

The FEHES Review Report (Tolonen 2008) provides a list for accessible sampling frames in each country. However, the list may not be complete. If no acceptable frame seems to be available, the national statistical institute or other national institutions, public or private, reg-

ularly carrying out national sample surveys in other fields should be consulted for assistance.

The target population is defined in Chapter 2 of Part A as all individuals of an eligible age, living in the country. Whenever the general Stage 2 frame does not cover all residents that should be eligible the various kinds of under-coverage should be explained and the size of the under-coverage estimated, preferably by sex and age. If for instance parts of the institutionalized population are not covered by the main frame for the survey, an overview of such institutions by category should be constructed. If feasible this should be done in such a way that this overview can be used as a supplementary Stage 1 frame for institutions although it will rarely be possible to take samples from the institutions.

3.3 Sampling design for Stage 1

3.3.1 Creating the PSUs

Whatever sampling frame for Stage 2 will be used, the sampling frame for Stage 1 (the PSUs) should be established approximately as follows.

Partition the geographical area of the country into a set of disjoint areas, the PSUs. Each PSU should be small enough to be served by one examination site and with acceptable travel distances to the site for all people living in the PSU or for field work team and mobile units to travel between the homes of all potential invitees. The PSUs should be areas for which statistics for total population sizes (number of persons) or the number of postal addresses or dwellings are accessible. What alternatives for PSUs are available may vary among countries, but small census tracts, municipalities, electoral districts and post code areas are examples. Most National Statistical Institutes in Europe have detailed population statistics by sex and age for all administrative units and sometimes also for smaller units defined for statistical purposes. For many countries this information is freely available on their websites and can be downloaded as excel files. If more detailed statistics is needed the statistical offices should be contacted.

From a statistical point of view it is desirable that PSUs, at least those within the same stratum, are statistically as similar as possible so that which PSUs are actually selected will affect the survey results as little as possible. As a PSU will have to be a contiguous area that will have to meet practical constraints there will always be limits to how similar it is possible to make them. They should however not be smaller than necessary to meet the practical demands since small PSUs will often tend to be internally more homogenous and therefore less similar to their neighbours. The sizes of the PSUs can vary within the same stratum, but not "too much".

3.3.2 Measure of size for the PSUs

A measure of size should be established for all PSUs. This will usually be the number of SSUs in the PSU according to the Stage 1 frame, people or household addresses, but if such up-to-date information is not immediately available, cruder measures of size, e.g. old census counts, should be used.

If the SSUs are people the size should be the number in eligible age living in the PSUs. If the distribution by sex and age is available this information should be taken into the file defining the Stage 1 frame. Age should be recorded by the groups that will be used for publication and comparison among countries, at least the age groups 25-34, 35-44, 45-54 and 55-64.

If the SSUs are households, dwellings or postal addresses, their numbers in the PSUs should be used as the measure of size. If the number of dwellings at each postal address is known, it will be better to use the number of dwellings than the number of addresses as the size measure of the PSU. If it is feasible to select dwellings directly rather than addresses the need for Stage 3 to select dwellings at multi-dwelling addresses can be avoided or reduced. In households with a large number of eligible individuals, it may be necessary to limit the number of participants. Techniques for doing this (Kish Grid, last birthday etc.) will not be discussed in this document. For an example, see the Health Survey of England (Craig 2008).

If neither a frame based on individuals nor addresses or dwellings is available for Stage 2, frames for further sampling must be established within the selected PSUs. An example of such a frame is the National Health and Nutrition Examination Survey (NHANES 1994).

3.3.3 Stratification of the PSUs

The PSUs should be stratified by grouping together relatively similar PSUs, e.g. urban PSUs versus rural PSUs, PSUs having similar age distribution by taking into account the social or demographic profile of the PSUs. Good stratification increases the precision of the survey estimates. Although the PSUs in a stratum do not need to be geographically contiguous, geography is also important. There is often interest in comparing regions within a country with respect to various health indicators. It is therefore desirable that these regions consist of complete strata. When considering how many strata to create, think about how many PSUs it is natural to select. Generally to facilitate variance estimation in a two-stage design, two PSUs should be selected per stratum. To be able to measure uncertainty in the estimates is important when comparing estimates from different countries or regions within a country. However, other considerations can justify selecting only one. Detailed stratification may reduce sampling variance but make unbiased estimation of the variance infeasible. Sometimes other considerations makes it is natural establish some small strata where selecting more than one PSU is difficult. PSUs that are very large in population can be strata alone (i.e. metropolitan areas). Very large PSUs can ei-

ther be cities where it can be seen as appropriate to sample in one stage or are those that are assigned a probability larger than one according to the formula. EHESsampling will automatically select PSUs that are too large compared to other PSUs in the same stratum with probability one. Such PSUs will be treated separately at Stage 2. As a basic rule, to be able to select two PSUs in a stratum it should contain at least four PSUs.

EHESsampling can calculate the number of PSUs to be sampled in each stratum and the anticipated costs doing the survey with this stratification. Using the software these calculations should be carried out for alternative stratifications as a tool to find the best ones, the one that gives the lowest variance for a given cost.

3.3.4 Sample sizes at Stage 1

For each stratum, the number m of PSUs to be selected at Stage 1 and the number p of SSUs to be invited within each PSU at Stage 2 can either be decided directly or be established based on cost-variance considerations. It will be demonstrated in this section and in Section 3.4.2 that if the PSUs are selected with Probability Proportional to Size (PPS), which is recommended, the Stage 2 sample size p should be the same within every sampled PSU in the same stratum.

How to calculate a cost-variance optimal value of m and p will be demonstrated below.

For a given stratum, let C_{PSU} be the average cost of sampling a PSU (i.e. setting up an extra site) in the stratum and let C_{SSU} be the average cost of inviting an (extra) SSU (person or dwelling) to the survey. Let m be the number of PSUs to be selected in the stratum and let n be the number of SSUs. A model for the expected variable survey cost in the stratum is then

$$C = C_{PSU}m + C_{SSU}n$$
(3.1)

Let Y be a survey variable. For a given stratum, let Y_{ij} be the value of this variable associated with SSU no. j in PSU no. i. Let N_i be the size (no. of main frame units) of PSU no. i. Let

$$\mu_i = \sum_{j=1}^{N_i} Y_{ij} / N_i = \text{ average of } Y \text{ within PSU } i \text{ and}$$

$$\sigma_i^2 = \sum_{j=1}^{N_i} (Y_{ij} - \mu_i)^2 / N_i = \text{ variance of } Y \text{ within PSU } i.$$

(3.2)

Let V_{Among} be the (weighted) *variance* (across PSUs) of the *within PSU averages* μ_i and let V_{Within} be the (weighted) *average (across PSUs) of the within* PSU *variances* σ_i^2, that is

$$N = \sum_{i \in S} N_i = \text{ total number of frame units in all PSUs in stratum S}$$

$$\mu = \sum_{i \in S} N_i \mu_i / N = \text{ average of } Y \text{ in stratum S}$$

$$V_{Among} = \sum_{i \in S} N_i (\mu_i - \mu)^2 / N = \text{variance among the PSUs of their averages}$$

$$V_{Within} = \sum_{i \in S} N_i \sigma_i^2 / N = \text{ average of the within PSU variances} \qquad (3.3)$$

We wish to sample *n* units within a stratum to estimate the average μ of *Y* in that stratum using a two-stage sample where every final unit has the same probability of being selected. When the PSUs are selected with probability proportional to size as described in Section 3.3.5 such an equal probability sample is obtained by allocating the sample size *n* equally with *p* = *n*/*m* units to each of the *m* sampled PSUs. The vari-

ance of the simple sample mean estimator $\hat{\mu} = \sum_{ij \in sample} Y_{ij} / n$ is then

$$Var(\hat{\mu}) = \frac{V_{Among}}{m} + \frac{V_{Within}}{n} \qquad (3.4)$$

The number of individuals *p* to be invited within a PSU and the number *m* of PSUs to be drawn to minimize the $Var(\hat{\mu})$ (given *n* and C) is given by the formulae

$$p_{opt} = \sqrt{\frac{V_{Within}}{V_{Among}} \frac{C_{PSU}}{C_{SSU}}}, \quad m = \frac{n}{p_{opt}} \qquad (3.5)$$

The value of p_{opt} obtained with formula will be different for different *Y*-variables. If formula is to be made operational, V_{Within} and V_{Among} must be calculated or estimated for some compromise *calculation variable*. Using a variable available in the sampling frame is best. Age is a recommended variable for this purpose since health in general depends strongly on age.

Note that p_{opt} does not depend on the total number of units (*n*) to be sampled in the stratum, but *m* does. Formula says that *if the sample size n of SSUs is to be increased within a stratum then this should be done by taking a larger sample of PSUs, not by selecting more SSUs within each PSU*. Note also that p_{opt} may be calculated larger than *n.* Since μ must be 2 or more, formula will only play a role if p_{opt} is calculated to be less than roughly 40 percent of *n*. *m* and *p* also need to be rounded to integers. This is done in the program EHESsampling.

Assessing the costs C_{PSU} and C_{SSU} is a part of the budgeting of the survey. The variances will differ among the variables and some compromised calculation values must be established to apply formula . Note

however that it is basically the ratios C_{PSU}/C_{SSU} and V_{Within}/V_{Among} that are needed to calculate p_{opt}. If nothing else can be assumed, C_{PSU}/C_{SSU} can be set equal for all strata. If information about age distribution within the PSUs comes with the sampling frame it is possible to establish values for V_{Within}/V_{Among} based on the age distribution within and across the PSUs. An example on calculation of V_{Within} and V_{Among} based on age distribution is given in section 3.6.1 and in the manual for EHESsampling.

Before calculating m, the number n of units (individuals or addresses) to be sampled in the stratum must be set. In an equal probability sample the total sample size should be allocated to the strata proportional to the number of units in the frame and this should be the basis for calculating m.

R-program EHESsampling does the calculations for m and p based on the given values for C_{PSU}/C_{SSU} and V_{Within}/V_{Among} and rounds the calculated sample sizes to integers in such a way that the total national sample size is maintained.

3.3.5 Inclusion probabilities for the PSUs

The sampling of Primary Sampling Units at the first stage should be done with Probability Proportional to Size (PPS-sampling). This means that the probability π_i for selecting PSU no. i in the stratum is

$$\pi_i = m\frac{N_i}{N}. \qquad (3.6)$$

Note that if m is large and there is significant variation in the sizes of the PSUs, formula (3.6) can assign the largest PSUs probabilities greater than one. This should be avoided if possible, and setting a minimum size for the PSUs aims at that. PSUs that are 'too large' in the sense that (3.6) produces a probability larger than 1, are automatically assigned $\pi_i = 1$. EHESsampling will select them and calculate n_i correctly among the rest. Specific examples will be provided in the manual for EHESsampling. EHESsampling will also at this stage calculate the Stage 2 inclusion probabilities and anticipated sample sizes to be used in each PSU if the PSU is selected at Stage 1. How the Stage 2 inclusion probabilities and sample sizes are calculated is described in Section 3.4.2.

3.3.6 Sampling

When all the preparations described have been completed, sampling can proceed using for example the R-program EHESsampling package. *PPS*-sampling is rather technical and there is a host of methods. EHES-

sampling uses an R-package called *sampling* developed by Yves Tillé and Alina Matei (2009).

Having chosen the desired number of PSUs to be selected, EHESsampling calculates the number of SSUs to be sampled at Stage 2 within each PSU if that PSU is being selected. If the sampling units are addresses or if age-sex stratification is not used (see Section 3.4.1), the sample sizes will be the same in all PSUs selected with probabilities less than one. There may be deviations of no more than one due to rounding. PSUs selected with probability equal to one will have larger sample sizes than the other PSUs in the same stratum. If age-sex stratification is used, the number of persons to be sampled to the same age-sex domain will vary among the PSUs in the same PSU-stratum. If EHES-sampling is used, the sample of PSUs along with the Stage 1 and 2 inclusion probabilities, PSU-size and anticipated sample sizes for Stage 2 is stored in an R data frame.

3.3.7 Distribution of PSUs over time

A national Health Examination Survey may be carried out over a long period of time, often a year, sometimes more. Teams may travel and visit each sampled PSU, one at a time. The examination site will be in operation for a limited period, from one day to a couple of weeks and then the team will move to another PSU. The order in which the PSUs are being visited is not indifferent. It is well known that there are seasonal variations in people's health caused by varying temperatures and weather conditions. If the teams starts operating in one part of the country, for instance in the southern part and then move gradually north to finish the survey in the northern part, the effects of season and geography on health variables which should be distinguishable in the data, will be confounded. It will be impossible to estimate them separately.

This should be considered when timing the survey. Ideally, a randomization of the order in which the sampled PSUs are visited is recommended but can be difficult to implement. If teams have to move across the country in a completely random order carrying a heavy cargo of equipment, the cost of travelling can be high both in terms of time and money.

For further discussion, see Part A, Chapter 6 of the EHES Manual.

3.4 Sampling design for Stage 2

Simple random sampling is proposed for sampling persons or addresses within each PSU selected at Stage 1. If age-sex stratification is used, simple random sampling should also be used within each age-sex domain in each selected PSU. However, age-sex stratification complicates the issue of sample sizes.

3.4.1 Stratification by age-sex domains

Results from national HESs will be compared across countries within age-sex domains. The four age domains 25-34, 35-44, 45-54 and 55-64 years will be crossed with sex to form eight domains. For this reason it is desirable to guarantee all eight domains a minimum sample size. With a minimum total sample size of 4 000 this means at least 500 persons in each domain. This can be obtained by a kind of stratification with respect to the eight domains where one sample is taken for each domain at Stage 2.

The eight age-sex domains will intersect the PSUs. Stratification with respect to domains that intersect the PSUs in a two-stage design is not common, but can be carried out so that every person in the same domain in the same stratum (or country) has the same inclusion probability. This is recommended when the total sample size for the survey is not large enough to by itself warrant the minimum sample size for each domain. When age-sex stratification is used, then the sample sizes will not be quite fixed. This is discussed in Section 3.4.2.

Age-sex stratification cannot be applied if the SSUs are addresses, dwellings or households.

3.4.2 Sample sizes at Stage 2 - with and without age-sex domains

We first consider the simplest case, without age-sex stratification. The goal is for all units within a stratum to have the same selection probability after two stages. This will be achieved if the Stage 2 sample sizes are calculated as:

$$n_i = n \frac{N_i}{N \pi_i} \tag{3.7}$$

If all PSUs are selected with probability less than one, ($\pi_i = mN_i/N$ as in (3.6)) the same sample sizes n_i at Stage 2 will all be equal:

$$n_i = \frac{n}{m} \tag{3.8}$$

for all i (PSUs) selected at Stage 1. The selection probability at Stage 2 is then

$$\varphi_i = \frac{n_i}{N_i} = \frac{n}{mN_i} \tag{3.9}$$

If the largest PSU, say PSU no. 1, is selected with probability $\pi_1 = 1$, then PSU no. 1 will have a different sample size that is calculated first:

$$n_1 = n\frac{N_1}{N}$$ (3.10)

where N_1 is the total number of sampling units in PSU no. 1. In this case $\varphi_1 = n/N$. The n_i s calculated are usually not integers. This does not matter. When sampling, each PSU gets a sample size which is n_i rounded either up or down in such a way that they sum to n and no PSU is statistically favoured.

If individuals are sampled and age-sex stratification is used at Stage 2, we recommend a solution where *every person in the same age-sex domain in the same stratum has the same selection probability after two stages whichever PSU the person belongs to in the stratum*. But the inclusion probabilities for the PSUs at Stage 1 have been set based on the total population in each PSU, not the population in each age-sex domain. Therefore, the total sample size for an age-sex domain within a stratum will depend somewhat on which PSUs have been selected at Stage 1 and so will the total sample size across all age-sex domains and PSUs in a stratum. Different domain definitions in different strata can be allowed, but different domain definitions in different PSUs in the same stratum are not allowed.

Starting at the top we define a *desired* sample size for an age-sex domain within a stratum, say n_d^* for domain *d in the actual stratum*. For instance, if you want a total sample size n for a stratum, and there are eight domains and you want to have approximately equal sample sizes for each domain you can choose $n_d^* = n/8$. Or if you want every member of domain *d* in your country to have the same probability of being selected, n_d^* should be set by allocating the nationally desired sample size proportional to the stratum size of domain *d*. In order to obtain equal probability sampling within the domain, the sample size for domain *d* within PSU no. *i,* if PSU no. *i* is being selected, must be

$$n_{id} = \frac{N_{id}n_d^*}{N_d\pi_i}$$ (3.11)

If all $\pi_i < 1$ this amounts to

$$n_{id} = \frac{n_d^*}{m}\frac{N}{N_d}\frac{N_{id}}{N_i}$$ (3.12)

Here N_d is the size of the age-sex domain *d* in the stratum, N_i is population size of PSU *i* across all domains and N_{id} is the population in domain *d* in PSU no. *i in the Stage1 frame*. The Stage 2 inclusion probabilities for all eligible individuals in domain *d* in PSU *i* is

$$\varphi_{id} = \frac{n_{id}}{N_{id}} \hspace{4cm} (3.13)$$

The sample sizes n_{id} are calculated for all PSUs before sampling at Stage 1. Their sum over the *selected* PSUs constitutes the *actual* sample size n_d in domain d in PSU no. i and depends on which PSUs have been selected. The actual sample size is therefore random and usually different from n_d^*, but is equal to n_d^* in expectation.

The variation of n_{id} should be kept as small as possible. Variation in age distribution and to some extent the sex distribution across the PSUs within strata will contribute to variability of the domain sample sizes. This is a strong argument *for considering age distribution when stratifying the PSUs*, whether age-sex stratification is being applied or not. We must expect to find considerable variation in age distribution across PSUs in all countries. Typically, recently established housing areas and areas with considerable immigration have a much younger population than districts where there is emigration and a declining population. It is most often the young population that moves.

The R-data frame produced by EHESsampling when taking the Stage 1 sample contains the anticipated sample sizes for Stage 2 in each selected PSU as shown in Table 3.3. The sampling variation in the domain sample sizes should be assessed for each age-sex domain, for instance with variance calculations. At this stage this has not been implemented in EHESsampling but might come in a future version. Studying the variation by simulating many Stage 1 samples is better than theoretical variance assessments and this option can also be implemented in a future version.

In practice in many countries the Stage 2 samples will often be selected later that the date when the Stage 1 frame was constructed and one PSU by one based on local registers. These local registers will show up PSU-sizes and domain-sizes that are different from the sizes N_i and N_{id} that we used when taking the Stage 1 sample and calculating φ_{id}. Correctly done, φ_{id} will be applied to the local registers to do the actual Stage 2 sampling. The real sample sizes will result from this process and they will be somewhat different from n_{id} in formulae (3.11) - (3.12). Never the less, the method will provide a sample where the selection probability $\pi_i \varphi_d$ is the *same for all individuals in the same age-sex domain in the same stratum* whichever PSU (i) the person belongs to in the stratum. This is the procedure offered in EHESsampling

The effect of age-sex stratification is illustrated in Table 3.1 with an example from the test population used in developing EHESsampling. It creates an overrepresentation of the smallest age-sex domains and an underrepresentation of the larger domains and thus a more even distribution among them in the sample, but not exactly equal sample size.

Table 3.1 Illustration of the effect of age-sex stratification on sample size

Domain	F25_34	F35_44	F45_54	F55_64	M25_34	M35_44	M45_54	M55_64	Total
Total sample	1318	1305	1317	1308	1338	1309	1295	1310	10500
Pct. of sample	12,55	12,43	12,54	12,46	12,74	12,47	12,33	12,48	100,00
Total in testpop	206329	238764	209751	187288	211359	248490	216226	187932	1706139
Pct. of testpop	12,09	13,99	12,29	10,98	12,39	14,56	12,67	11,02	100,00

A simpler procedure for age-sex domain stratification at Stage 2 can be selected.

1. Decide the total sample size at Stage 2 for every (selected) PSU, say n_i = 200
2. Divide this size by the number of domains, e.g. n_{id} = 200/8 = 25.

And take 25 as the sample size for all domains in the PSUs. (3.13) will still be valid. The stratum domain sample size n_d will be exactly as desired. However, this method will *not* yield $\pi_i \varphi_{id}$ the same for every individual in the same age-sex domain in the same stratum whichever PSU (*i*) the person belongs to in the stratum.

3.4.3 When the Stage 1 frame is approximate

The sampling frame used to establish the Stage 1 frame and to do the calculations described in Section 3.4.2 may be approximate in relation to the actual sampling process. The selection of the PSUs will have to take place well before the survey is carried and may be based on population statistics that are not up-to-date. A Stage 2 sampling frame with the desired SSUs may also exist only locally and not centralised at national level. Furthermore, since the survey will have to take place over a year or more, those selected at the beginning of the survey period may have died or moved from the PSU at the time when the survey team establishes a clinic there. Therefore, it is desirable to take the Stage 2 sample for a PSU as close as possible to the time when the PSU will be visited. This may mean that the Stage 2 samples have to be taken at different times for different PSUs. When considering eligibility and age-sex domains, the age of a person included in the Stage 2 frame should then be taken as the *age at the middle of the data collection period in the actual PSU*. Then the sizes of the PSUs (N_i) and the age-sex domains (N_{id}) may have changed slightly and the anticipated Stage 2 sample sizes calculated in Section 3.4.2 will be adjusted somewhat according to that.

When adjusting the sample sizes we wish to maintain the Stage 2 inclusion probabilities φ_i (or φ_{id}) and over all inclusion probabilities $\pi_i \varphi_i$ (or $\pi_i \varphi_{id}$) calculated in Section 3.4.2. This means that the sample sizes for Stage 2 will have to be recalculated based on new population counts. The Stage 2 sampling procedure described in Section 3.4.4 will automatically handle this.

3.4.4 Taking the Stage 2 sample

To prepare for sampling at Stage 2, select from the main sampling frame all individuals or addresses that belong to the PSUs selected in Stage 1. The sampling procedure described below is documentation for the interested reader. Other readers can rely on EHESsampling which has a module that does it all.

1. In the data frame consisting of all Stage 2 sampling units in the selected PSUs, associate the Stage 2 inclusion probabilities φ_i (φ_{id}) to each record.
2. For each record, generate a random number u between 0 and 1.
3. Sort the data frame by PSU by age-sex domain by u.
4. Create a new variable a by aggregating the φ_i (φ_{id}) successively up to the previous record in the file and $b = a + \varphi_i$.
5. Generate a new random number r between 0 and 1.
6. Find the record in the frame for which $a \le r < b$. This record is selected.
7. Set $r = r + 1$.
8. Go to step 6 and repeat the procedure until you are through all records in the file.

This provides equal probability random samples in all domains by PSUs. If the calculated sample size in a selected PSU is say 23.84, then the actual sample size by this algorithm will be 24 with probability 0.84 and 23 with probability 0.16.

3.5 When using address frames

3.5.1 Multi-dwelling houses

If the main sampling frame is addresses and each dwelling at multiple dwelling addresses cannot be identified in the frame, a third sampling stage may be necessary. It is not desirable to include all dwellings at an address with many dwellings.

If the *number* of dwellings at an address (k) is known, copy the record for the address so that there are k records for the address in the frame. Take the sample as described previously and apply a unique rule for assigning a physical dwelling to the selected records. The advantage of this is that all dwellings at such multiple dwelling addresses will still

have the same probability of being selected as single-address dwellings.

If the number of dwellings at an address is not known, the selection of dwellings should occur when the address is visited for the first time. The dwellings must be mapped and the number of dwellings to be selected must be decided. Sampling dwellings at an address can be seen as a third stage in the design. If the addresses have been sampled with equal probability, taking a sample of the dwellings at multi-dwelling addresses means that such dwellings will have lower inclusion probabilities than single-address dwellings. This is a disadvantage which will have to be corrected by weighting. To reduce this disadvantage at least two dwellings should be sampled at each such address. On the other hand, sampling many dwellings at multi-dwelling addresses will increase the total number of dwellings and people in the sample.

3.5.2 Selection of individuals within a dwelling

When using address frames, the participant invitation must take place when visiting the address and the selected dwelling for the first time. In both cases, the selection must be random. Random selection of individuals within a dwelling can be seen as a third (or in some cases fourth) stage of sampling and will affect the actual sample sizes, inclusion probabilities and the sampling weights to be used later in estimation.

Basically, everyone in the core age group (25-64 years) living in a selected dwelling should be invited to the survey. This gives every person the same probability of being selected, independent of household size. If the number of eligible people in the dwelling is very high, a maximum should be set. Defining this limit is a national decision, but should not be less than three. To select participants, all eligible individuals in the dwelling must be listed and a random sample must be taken from that list. The selection probabilities for people living in such dwellings will be lower than for those living in "take all" dwellings by a factor equal to the fraction of eligible persons selected in the dwelling. This must be corrected for by proper weighting at the estimation stage. For selection of individuals within a dwelling one can use a (modified) Kish grid. For references to Kish grid techniques see for instance Kish, (1949, 1965) and Nemeth (2001, 2003).

3.5.3 Other situations

When neither of the kinds of sampling frames mentioned in item 1 or 2 in section 3.2 are available, it should be possible to carry out Stage 1 in the sampling design much like when individual or postal-address frames are available. But small area population counts from censuses or other sources must be available at a suitable level. Detailed discussions of such cases will not be done here, but will be taken with the countries which it concerns.

3.6 Documentation and data management

An overview and details for data management are described in Part A, Chapter 12 of the EHES Manual. This section will consider the documentation of the sampling design and the samples produced by that design.

3.6.1 Reporting the sampling at Stage 1

The documentation for Stage 1 sampling must describe the sampling frame for Stage 1, which kind of units are being used for PSUs, how many they are, their stratification and how the PSUs have been selected within each stratum. The documentation must contain two files/tables with a minimum set of columns described below. If EHESsampling is being used for organizing the sampling and selecting the sample, it will produce R data frames with all the information that we ask for. The preferred format when submitting the files to EHES RC is semicolon separated ascii (CSV) text file. In each file, the first row should be for the variable names, also separated by semicolons. The main features of the files are described below.

A. Stratification file

A file that describes the stratification. This file must have a name with the format EHES_*CC*_*SC*_stratification. Here *CC* represents the EU's two-letter *Country Code* and *SC* represents a two digit *Survey Code* that identifies different EHES-surveys within the same country. See chapter 12.2. The file must contain one row for each stratum. Below is the list of variables for this file. The variable names are typed in **ARIAL (bold)**.

1. **COUNTRY**. Character (2) *Country Code CC*.
2. **SURVEY**. Character (2). The *Survey Code SC*.
3. **STRATUM_ID**. Character (max 3). A stratum identifier (code).
4. **STRATUM_NAME**. Character (max 20). Common name of stratum
5. **STRATUM_SIZE**. Integer. The size of the stratum. The total number of SSUs, (N)
6. **DOMAINS**. Integer. The number of age-sex strata in Stage 2 sampling. = 1 if no age-sex stratification is used.
7. **ST1_ANT_SSU**. Decimal (2). The anticipated number of SSUs to be selected within the stratum (n)
8. **ST1_NO_PSU**. Integer. The number of PSUs in the stratum (M_{PSU})
9. **ST1_SEL_PSU**. Integer. The number of PSUs to be selected in the stratum (m)

10. **ST1_CV.** = 1 if **ST1_SELNO_PSU** has been calculated using cost-variance optimization (section 3.3.4). = 2 otherwise.

The following items are only relevant if **ST1_CV** = 1.

11. **ST1_CPSU.** Integer. The average cost of establishing a PSU in the stratum (C_{PSU})

12. **ST1_CSSU.** Integer. The average cost of inviting SSU in the stratum (C_{SSU})

13. **ST1_WITHIN.** Decimal (4). The average within PSU variance of the calculation variable (V_{Within})

14. **ST1_AMONG.** Decimal (4). The variance of the PSU means for the calculation variable (V_{Among})

15. **ST1_COST.** (Optional). Integer. The total cost of carrying out the survey in the stratum as calculated by formula ($Cost$)

Table 3.2 shows an example stratification file. The correspondence between the variable names in the formulae and the variable names in the file is shown in the two-line heading. The variable **DOMAINS** has been set to 8 for all strata indicating that age-sex stratification with eight domains is used I all strata. The variable **ST1_SEL_SSU** (n) has in this example been calculated based on a proportional allocation of 9000 sampled individuals but with sample sizes less than 400 adjusted up to a minimum of 400 persons in each stratum. Notice that n has been calculated by an allocation formula which usually does not produce an integer result and is therefore given with decimals. In EHESsampling the rounding to an integer will take place in the sampling process at Stage 2 (see Section 3.4.4). **ST1_VWITHIN** and **ST1_VAMONG** (V_{Within} and V_{Among}) have been based on age coded with 1 = '25-34 years', 2 = '35-44 years', 3 = '45-54 years' and 4 = 55-64 years' and calculated using formula (3.3). This is sufficient accuracy for the purpose although 'Age' could have been used more directly. The values of **ST1_CPSU** and **ST1_CSSU** (C_{PSU} and C_{SSU}) in this example are not real costs, but 'raw guesses' made up for testing purposes. The variable **ST1_COST** ($Cost$) is the total cost of carrying out the survey in the actual stratum calculated according to formula (3.1) in Section 3.3.4 with the values of m, n, C_{PSU} and C_{SSU} given in the table. The variable **STRATUM_NAME** provides a common name for the stratum in addition to its code.

As already mentioned EHESsampling produces a table with the variables that we ask you to report.
Comment: In stratum 03 in the example the value for V_{Among} is large compared to the other strata. This results in a low value for p_{opt} and a high value for m. The high value of V_{Among} and the slightly low value of V_{Within} express a large variation of average age among the PSUs in that stratum which may be typical for cities. In such cases the precision of the survey would benefit from splitting the stratum in two strata, one stratum for the PSUs with average age less than the median (or mean) and one for the PSUs with average age above the median for the PSUs. This would also result in having to select a smaller number of PSUs and a lower cost for that stratum. An alternative is to combine PSUs to larger and less homogenous PSUs as long as they do not become too large to be suitable for the survey.

A stratum file will have to be produced before the actual Stage 1 sample is taken. A file that is used as the input to EHESsampling can have ready made columns for sample sizes (*n* and *m*), possibly two or three alternatives in which EHESsampling can do calculations. Or it can contain no columns for *n* and *m* and let EHESsampling calculate them. However, the stratification file to be reported should only contain the sample sizes for the design actually used.

B. Primary Sampling Unit (PSU) file

The PSU-file to be reported is a file that describes the *selected* PSUs only, the Stage 1 sample. This file is described below. But before taking the Stage 1 sample EHESsampling will establish a PSU file similar to the one we ask you to report but with all PSUs in the frame. Input for establishing this file is a file with the more basic variables from which the remaining new variables are calculated. The input file is described in the EHESsampling manual. The Stage 1 sample will be a sample of PSUs from this file. The file name should have the format EHES_*CC*_*SC*_PSUSAMPLE where *CC* and *SC* follow the same standard as for the stratum file A. *If age-sex stratification will not be used at Stage 2 the PSU-file must contain one record for each selected PSU. If age-sex stratification is to be used the file should contain one row for each PSU by age-sex domain in each selected PSU.* Each row should contain the variables

Table 3.2 Example Stratification File (EHES_NO_CC_stratification)

				N		*n*	M_{PSU}	*m*		C_{PSU}	C_{SSU}	V_{Within}	V_{Among}	*Cost*
COUNTRY	SURVEY	STRATUM_ID	STRATUM_NAME	STRATUM_SIZE	DOMAINS	ST1_ANT_SSU	ST1_NO_PSU	ST1_SEL_PSU	ST1_CV	ST1_CPSU	ST1_CSSU	ST1_VWITHIN	ST1_VAMONG	ST1_COST
NO	01	01	Østfold	72970	8	400.00	43	2	1	10000	50	1.1531	0.0068	40000
NO	01	02	Akershus	259795	8	1370.44	99	9	1	8000	45	1.0905	0.0091	133670
NO	01	03	Oslo	175955	8	928.17	139	22	1	8500	60	1.0936	0.0908	242690
NO	01	04	Hedmark	65801	8	400.00	34	3	1	7000	55	1.1644	0.0060	43000
NO	01	05	Oppland	68775	8	400.00	42	2	1	8000	50	1.1563	0.0054	36000
NO	01	06	Buskerud	101992	8	538.01	56	4	1	9000	60	1.1610	0.0104	68281
NO	01	07	Vestfold	70045	8	400.00	43	3	1	9500	70	1.1313	0.0078	56500
NO	01	08	Telemark	32045	8	400.00	18	3	1	8500	65	1.1592	0.0070	51500
NO	01	09	Aust-Agder	35719	8	400.00	21	3	1	7500	45	1.1788	0.0100	40500
NO	01	10	Vest-Agder	73670	8	400.00	37	3	1	8000	50	1.1594	0.0099	44000
NO	01	11	Rgaland	182525	8	962.83	89	10	1	10000	55	1.1446	0.0235	152956
NO	01	12	Hordaland	169316	8	893.15	108	10	1	9500	40	1.1529	0.0358	130726
NO	01	14	Sogn og Fjordane	26729	8	400.00	12	2	1	8000	35	1.1471	0.0074	30000
NO	01	15	Møre og Romsdal	63492	8	400.00	39	3	1	8500	50	1.1636	0.0099	45500
NO	01	16	Sør-Trøndelag	112555	8	593.74	59	9	1	7500	55	1.1729	0.0347	100155
NO	01	17	Nord-Trøndelag	47782	8	400.00	22	2	1	6500	50	1.1630	0.0038	33000
NO	01	18	Nordland	63237	8	400.00	35	2	1	8000	45	1.1505	0.0042	34000
NO	01	19	Troms	58134	8	400.00	33	4	1	9000	60	1.1535	0.0183	60000
NO	01	20	Finnmark	25602	8	400.00	15	3	1	9500	75	1.1591	0.0072	58500

1. **COUNTRY**. Character (2). *Country Code CC*.

2. **SURVEY.** Character (2). The *Survey Code SC*

3. **STRATUM_ID.** Character (max 3). A stratum identifier (code).

4. **PSU_SN**. Character (max 4). A PSU serial number (maximum four digits) which *replaces* the real PSU ID (e.g. postcode, municipality code etc.) that has to be used nationally to identify the PSU for data collection. The purpose of the PSU serial number is to tell which individuals or households belong to the same PSU. This is important for proper analysis of sampling variance. It will not be necessary or even desirable for the Reference Centre to know which real PSU is represented by a PSU serial number and for confidentiality reasons that information should not be transferred. We recommend that the serial numbers run across strata since this will distinguish PSUs without using the stratum variable. A link between the PSU serial number and the real PSU ID should be maintained by the national survey organizer only.

5. **PSU_SIZE**. Integer. The size of the PSU (N_i)

6. **ST1_PROB**. Decimal (4). The Stage 1 inclusion probability (π_i) used in sampling

7. **ST2_PROB.** Decimal (4). The Stage 2 inclusion probability or for the PSU (φ_i) or domain (φ_{id}) (if **DOMAINS** > 1 in the stratification file). Notice that if age-sex stratification is used this probability will be different for different age-sex domains.

8. **ST2_ANT_SSU**. Decimal (4). Optional. Anticipated sample size within the PSU or age-sex domain (n_i or n_{id})

   If age-sex stratification is used (**DOMAINS** > 1 in the stratification file) in at least one stratum it should also contain

9. **DOMAIN_ID**. Character (max 10). A domain identifier specifying the age-sex domain for the record

10. **DOMAIN_SIZE_PSU**. Integer. The number of people in each age-sex domain in the PSU (N_{id}).

11. **DOMAIN_SIZE_STR**. Integer. The number of people in each age-sex domain within the stratum (N_d).

Table 3.3 presents an excerpt from an example PSU file without age-sex stratification and Table 3.4 present an excerpt from an example with age-sex stratification. In these files, postcode is used as the PSU identifier.

In Table 3.3 **ST1_PROB** (π_i) is calculated from formula (3.6) and **ST2_ANT_SSU** (n) is calculated from formula (3.8), Section 3.4.2 and **ST2_ANT_SSU** is the same for all PSUs in the same stratum and will be rounded to an integer in the Stage 2 sampling. **ST2_PROB** (φ_i) is calculated by formula (3.9) and the final selection probabilities after

two stages $\pi_i\varphi_i$ will be the same for all PSUs (and SSUs) in the same stratum.

Table 3.3 Excerpt from a Primary Sampling Unit file without age-sex domains

				N_i	π_i	φ_i	n_i
COUN-TRY	SURVEY	STRA-TUM	PSU_SN	PSU_SIZE	ST1_PROB	ST2_PROB	ST2_ANT_SSU
NO	01	01	098	1886	0.0517	0.1060	200.00
NO	01	01	131	1433	0.0393	0.1396	200.00
NO	01	02	001	2549	0.0883	0.0597	152.27
NO	01	02	007	5681	0.1968	0.0268	152.27
NO	01	02	013	2767	0.0959	0.0550	152.27
NO	01	02	017	1958	0.0678	0.0778	152.27
NO	01	02	119	2364	0.0819	0.0644	152.27
NO	01	02	026	2999	0.1039	0.0508	152.27
NO	01	02	054	7316	0.2534	0.0208	152.27
NO	01	02	067	2869	0.0994	0.0531	152.27
NO	01	04	018	6983	0.3184	0.0191	133.33
NO	01	04	037	2099	0.0957	0.0635	133.33
NO	01	04	083	1422	0.0648	0.0938	133.33
NO	01	05	037	1244	0.0362	0.1608	200.00
NO	01	05	081	2217	0.0645	0.0902	200.00
NO	01	06	105	1426	0.0559	0.0943	134.50
NO	01	07	022	1700	0.0728	0.0784	133.33
NO	01	07	024	1059	0.0454	0.1259	133.33
NO	01	07	067	1414	0.0606	0.0943	133.33

In Table 3.4 **ST2_DOMAINS** = 8 and each PSU is represented by eight rows, one row for each age-sex domain, labelled by the variable **ST2_DOMAIN_ID**. **ST1_PROB** (π_i) is calculated in the same way as in Table 3.3. **ST2_ANT_SSU** (n_{id}) is calculated from (3.12) and varies over the domains in the same PSUs, but their sums over all domains are the same as in Table 3.3. **ST2_PROB** (φ_{id}) has been calculated using for-

mula (3.13). The final selection probabilities after two stages $\pi_i\varphi_{id}$ will be the same for the same age-sex domain in all PSUs (and SSUs) in the same stratum but differs across domains.

Countries that do age-sex stratification by simply taking the same number of persons in each age-sex domain in each PSU (see end of Section 3.4.2) should report that number for **ST2_ANT_SSU. ST2_PROB** (φ_{id}) should still be calculated using formula (3.13), but since **ST2_ANT_SSU** (n_{id}) will be different the final selection probabilities after two stages

$\pi_i\varphi_{id}$ will *not* be the same for the same age-sex domain in all PSUs (and SSUs) in the same stratum.'

Table 3.4 Example Stratification File (EHES_NO_CC_stratification)

COUNTRY	SURVEY	STRATUM	PSU_SN	PSU_SIZE N_i	ST1_PROB π_i	DOMAIN_ID	DOMAIN_SIZE_PSU N_{id}	DOMAIN_SIZE_STR N_d	ST2_PROB φ_{id}	ST2_ANT_SSU n_{id}
NO	01	02	103	3957	0.1371	M25_34	377	26141	0.0478	18.02
NO	01	02	103	3957	0.1371	M35_44	663	39975	0.0313	20.73
NO	01	02	103	3957	0.1371	M45_54	601	35218	0.0355	21.33
NO	01	02	103	3957	0.1371	M55_64	337	28111	0.0445	14.98
NO	01	02	103	3957	0.1371	F25_34	386	27438	0.0455	17.58
NO	01	02	103	3957	0.1371	F35_44	704	40418	0.0309	21.77
NO	01	02	103	3957	0.1371	F45_54	554	33824	0.0369	20.47
NO	01	02	103	3957	0.1371	F55_64	335	28670	0.0436	14.60
NO	01	02	029	4133	0.1432	M25_34	304	26141	0.0458	13.91
NO	01	02	029	4133	0.1432	M35_44	690	39975	0.0299	20.65
NO	01	02	029	4133	0.1432	M45_54	610	35218	0.0340	20.72
NO	01	02	029	4133	0.1432	M55_64	437	28111	0.0426	18.60
NO	01	02	029	4133	0.1432	F25_34	368	27438	0.0436	16.05
NO	01	02	029	4133	0.1432	F35_44	685	40418	0.0296	20.28
NO	01	02	029	4133	0.1432	F45_54	629	33824	0.0354	22.25
NO	01	02	029	4133	0.1432	F55_64	410	28670	0.0417	17.11
NO	01	02	052	1698	0.0588	M25_34	112	26141	0.1114	12.48
NO	01	02	052	1698	0.0588	M35_44	249	39975	0.0729	18.14
NO	01	02	052	1698	0.0588	M45_54	255	35218	0.0827	21.09
NO	01	02	052	1698	0.0588	M55_64	218	28111	0.1036	22.58
NO	01	02	052	1698	0.0588	F25_34	176	27438	0.1061	18.68
NO	01	02	052	1698	0.0588	F35_44	250	40418	0.0721	18.01
NO	01	02	052	1698	0.0588	F45_54	220	33824	0.0861	18.94
NO	01	02	052	1698	0.0588	F55_64	218	28670	0.1016	22.14

3.6.2 Reporting the sampling at Stage 2

The sample resulting from Stage 2 must be reported to the EHES RC in full. *All sampled units must be included in the file, even if they were later found to be ineligible to the sample or they did not participate in the survey*. The filename format should be EHES_*CC_SC*_st2sample. There should be *no direct identifiers* such as ID number, names or addresses on the file. The file should contain

1. **COUNTRY**. Character (2). The two-character *Country Code CC*.

2. **SURVEY.** Character (2). The two digit *Survey Code SC*.

3. **STRATUM_ID.** Character (max 3). A stratum identifier, max 3 characters.

4. **PSU_SN.** Character (max 4). A PSU serial number (maximum four digits), the same as in the PSU-file.

5. **SERIAL.** Character (max 12). Serial number that uniquely identifies a person within the survey. When the SSUs are individuals, the number can be assigned immediately after the sample has been selected. EHESsampling provides option for this. If households are used as SSUs the serial number must be assigned when the household is visited. Must not contain information that identifies the person in the population. See section 12.2. Assigned after visit of household if **HOUSEHOLD_UNIT** = 1.

6. **ST2_DOMAIN_ID.** Character (max 10). A domain identifier specifying the age-sex domain for the record. Only relevant if **DOMAINS** > 1 in the stratum file. Then equal to **DOMAIN_ID** in that file.

7. **ST2_SEL_SSU.** (Integer). Number of SSUs actually selected within the PSU or domain. Must be calculated when the Stage 2 sampling has taken place. All who were selected should be counted here, also those who were later found to be not eligible to the sample and those who did not respond.

8. **HOUSEHOLD_UNIT.** = 1 if addresses/households/dwellings are used as SSUs. = 2 otherwise

9. **HOUSEHOLD_SN.** Character (max 5). Relevant only if **HOUSEHOLD_UNIT** = 1. An address or *Household Serial Number* (*HSN*) with maximum five digits (code "88888" on all records if addresses or households are *not* used as sampling units). The number can be assigned within or across the PSU serial numbers.

10. **ST3_SAMPLING.** Only relevant if **HOUSEHOLD_UNIT** = 1. =1 if there is probability sampling within households. = 2 otherwise.

11. **ST3_PROB.** Decimal (4). Stage 3 inclusion probabilities. If **ST3_SAMPLING** = 1 then the stage 3 incusion probability. Otherwise ST3_PROB= 1.0000.

12. **ALL_PROB.** The overall inclusion probability. **= ST1_PROB * ST2_PROB * ST3_PROB.**

Table 3.5 Excerpt from a Stage 2 sample (with age-sex stratification)

COUNTRY	SURVEY	STRATUM	PSU_SN	SERIAL	HOUSEHOLD_UNIT	DOMAIN_ID	ST2_SEL_SSU	HOUSEHOLD_SN	ST3_SAMPLING	ST3_PROB	ALL_PROB $\pi_i \varphi_{id}$	SAMPLING_WEIGHT $1/\pi_i \varphi_{id}$
NO	01	03	004	00261	2	M25_34	4	88888	2	1.0000	0.0036	277.189
NO	01	03	004	00262	2	M25_44	4	88888	2	1.0000	0.0036	277.189
NO	01	03	004	00263	2	M25_44	4	88888	2	1.0000	0.0036	277.189
NO	01	03	004	00264	2	M25_34	4	88888	2	1.0000	0.0036	277.189
NO	01	03	004	00265	2	M35_44	5	88888	2	1.0000	0.0045	221.691
NO	01	03	004	00266	2	M35_44	5	88888	2	1.0000	0.0045	221.691
NO	01	03	004	00267	2	M35_44	5	88888	2	1.0000	0.0045	221.691
NO	01	03	004	00268	2	M35_44	5	88888	2	1.0000	0.0045	221.691
NO	01	03	004	00269	2	M35_44	5	88888	2	1.0000	0.0045	221.691
NO	01	03	004	00270	2	M45_54	3	88888	2	1.0000	0.0065	154.031
NO	01	03	004	00271	2	M45_54	3	88888	2	1.0000	0.0065	154.031
NO	01	03	004	00272	2	M45_54	3	88888	2	1.0000	0.0065	154.031
NO	01	03	004	00273	2	M55_64	4	88888	2	1.0000	0.0083	120.753
NO	01	03	004	00274	2	M55_64	4	88888	2	1.0000	0.0083	120.753
NO	01	03	004	00275	2	M55_64	4	88888	2	1.0000	0.0083	120.753
NO	01	03	004	00276	2	M55_64	4	88888	2	1.0000	0.0083	120.753
NO	01	03	004	00277	2	F25_34	6	88888	2	1.0000	0.0036	277.189
NO	01	03	004	00278	2	F25_34	6	88888	2	1.0000	0.0036	277.189
NO	01	03	004	00279	2	F25_34	6	88888	2	1.0000	0.0036	277.189
NO	01	03	004	00280	2	F25_34	6	88888	2	1.0000	0.0036	277.189
NO	01	03	004	00281	2	F25_34	6	88888	2	1.0000	0.0036	277.189
NO	01	03	004	00282	2	F25_34	6	88888	2	1.0000	0.0036	277.189
NO	01	03	004	00283	2	F35_44	4	88888	2	1.0000	0.0051	196.920
NO	01	03	004	00284	2	F35_44	4	88888	2	1.0000	0.0051	196.920
NO	01	03	004	00285	2	F35_44	4	88888	2	1.0000	0.0051	196.920
NO	01	03	004	00286	2	F35_44	4	88888	2	1.0000	0.0051	196.920
NO	01	03	004	00287	2	F45_54	3	88888	2	1.0000	0.0068	146.550

13. SAMPLING_WEIGHT. Decimal (4). **= 1/ ALL_PROB**.

With reference to Part A, Section 12.2, a SERIAL NUMBER must be given to everybody selected to the sample (i.e. not only for example those eventually found eligible or to the survey participants). This serial number must be unique to every person within the survey. Whenever the sampling units are individuals this serial number should be assigned immediately after sampling.

However, when addresses or households are used as sampling units, *SN* can be completed only at the stage when the household is visited and all eligible subjects at the address or in the household have been mapped.

3.7 Some common designs – a discussion

Two stage sampling is a complicated matter. There are some common ways of doing two-stage sampling that we have not recommended. One of these is to sample the PSUs with equal probability within each stratum at Stage 1, perhaps only one or two PSUs. This will produce a sample with many more of the smaller PSUs compared to the large ones than a probability proportional to size (PPS) design. Another strategy is to sample the same proportion of SSUs in each selected PSU at Stage 2. Depending on which combination of strategy for Stage 1 and Stage 2 one chooses one will get equal or unequal selection probabilities after two stages, over or under representation of SSUs in small PSUs versus large ones or fixed or random total sample size. Random sample sizes results in an unpredictable number of invited participants. This is disadvantageous both from a statistical, cost and administrative point of view since both will depend heavily on the sample size. The four combinations and their respective advantages and disadvantages are depicted in Table 3.6. Notice the only design recommended in this chapter requires both fixed sample sizes and equal selection probabilities for every secondary sampling unit (SSU) after two stages. The table assumes that the Stage 1 design provides a fixed number of PSUs in the Stage 1 sample.

Table 3.6 Combinations of sampling strategies for Stage1 and Stage2

Combinations of sampling strategies for Stage 1 and Stage 2		
Stage 2	**Stage 1**	
	A. Euqal probabilities selection	**B. Probabilities pro-portions to size**
1. Same sampling pro-portions in all PSUs	Equal probabilities for all people/-households after two stages within stratum. Small samples in small PSUs. Large samples in large PSUs. Final total sample size unpredictable and depends on which PSUs are selected at stage 1.	Large samples in large PSUs and small samples in small PSUs in the same stra-tum. Unequal selection probabilities after two stages: People/households in large PSUs over represented Final total sample size unpredictable and depends on which PSUs are selected at stage 1.
2. Same sample sizez in all PSUs	Fixed sample sizes after two stages. Smaller stage 2 prob-abilities (sampling proportions) for people/-households in large PSUs than small PSUs resulting in unequal selection prob-abilities after two stages: People/households in large PSUs are under represented compared to people/households in small PSUs.	Fixed sample size after two stages. Equal selection proba-bilities after two stages. **RECOMMENDED DE-SIGN**

Combination of stratification of PSUs by size and strategy A2 is com-mon. A strategy seen in some countries has been to first stratify the PSUs by regions and within each region by three sizes. If the same number of PSUs is selected in each size-stratum, people and house-holds in large PSUs will be underrepresented in the total sample. It is possible to compensate for this by selecting more PSUs in the strata for large PSUs, but it will be difficult to establish exact equal probability samples that way.

Selecting only one or two PSUs per stratum may save the costs of establishing a large number of examination clinics, but may lead to larger sampling variances than selecting a larger number of PSUs and less people within each of them, in particular if the PSUs are not very

similar with respect to relevant characteristics (V_{Among} is large). To find a good balance between cost and variances is the purpose of allocation formula .

On the other hand, if it is possible to make a detailed stratification with homogenous strata (V_{Among} is small), selecting only two PSUs per stratum may be optimal. This is shown in Table 3.2 in Section 3.6. Even selecting one PSU per stratum may be cost-variance optimal. But this will render unbiased estimation of the Stage 1 component of the variances infeasible in these strata since no variation among the PSUs will then be visible in the data. For this reason EHESsampling always selects at least two PSUs per stratum.

References

- Craig R, Shelton N, (eds). Health Survey for England 2007. Volume 2: Methodology and Documentation. Leeds: The NHS Information Center; 2008.
- Axelson M, Bihler W, Djerf K, et al (prepared by) European Health Interview Survey. The Task Force III report on sampling issues. 2009. Available at https://circabc.europa.eu/sd/d/4f89b382-b3e1-4884-94e1-ca9f2e2338b6/Task%20Force%20III%20report%20on%20sampling%20issues%20up%20to%20date%20November%202009.pdf. Accessed 22 January 2013.
- Jentoft S, Heldal J. RcmdrPlugin.EHESsampling v. 2.0. [Software Package]. Comprehensive R Archive Network (CRAN); 2011.
- Kish L. A procedure for objective respondent selection within the houshold. Journal of the American Statistical Association. 1949 Sep;44(247):380-7.
- Kish L. Survey Sampling. New York: Wiley; 1965.
- Nemeth R. Respondent selection within the houshold: A modification of the Kish Grid. 2001; Available from: http://www.math.uni-klu.ac.at/stat/Tagungen/Ossiach/Nemeth.pdf.
- Nemeth R. Sampling design of Health Surveys: Household as the sampling unit. Luxembourg Income Study Working Paper Series. 2004; (358). Available from http://www.lisdatacenter.org/wps/liswps/358.pdf
- NHANES. National Health and Nutrition Examination Survey (NHANES). Vital and health statistics. Plan and operation of the third National Health Examination Survey 1988-94. 1994.
- Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National Health and Nutrition Examination Survey: Sample design, 2011–2014. National Center for Health Statistics. Vital Health Stat 2(162). 2014.
- Tolonen H, Koponen P, Aromaa A, Conti S, Graff-Iversen S, Grøtvedt L, et al. (eds.) Reveiw of Health Examination Surveys in Europe. (2008) Publications of the National Public Health Institute. Helsinki. B18/2008. Available from http://urn.fi/URN:ISBN:978-951-740-843-1. Accessed on 22 January 2013

Additional literature on sampling

- Tolonen H, Koponen P, Aromaa A, Conti S, et al (eds.). Recommendations for Health Examination Surveys in Europe. (2008) Publications of the Na-

tional Public Health Institute. Helsinki. B21/2008. Available from:http://urn.fi/URN:NBN:fi-fe201204193358. Accessed on 22 January 2013.

- Lehtonen R, Pahkinen E. Practical Methods for Design and Analysis of Complex Surveys. New York: Wiley; 2004.

- DEGS. Operationshandbuch. Berlin: Robert Koch Institut; 2010.

- Rao PSRS. Sampling Methodologies with Applications. Chapman & Hall; 2000

- Särndal CE, Swensson B, Wretman J. Model assisted Survey Sampling. New York: Springer-Verlag; 1992. (Springer Series in Statistics).

- Tillé Y. Sampling Algorithms. New York: Springer; 2006. (Springer Seriers in Statistics).

- Thompson SK. Sampling. 2nd ed. Wiley; 2002

- Valliant R, Dorfman AH, Royall RM. Finite Population Sampling and Inference – A Prediction Approach. Wiley; 2000

4. Legal and ethical aspects

Susanna Conti[1], Mark Kanieff[1], Grazia Rago[1], Laura Paalanen[2], Päivikki Koponen[2]

[1] Istituto Superiore di Sanità (ISS), Rome, Italy
[2] National Institute for Health and Welfare (THL), Helsinki, Finland

PART A

In any research involving humans, ethical conduct is a fundamental concern. This means that the research must be performed so that participants are protected not only from risks to their physical and mental health but also from risks to their privacy and from receiving misinformation. Although performing a HES does not pose any serious risk to health, the safeguarding of privacy and acquiring informed consent are crucial ethical aspects. In this chapter, we describe a series of recommendations related to the legal and ethical aspects of performing a HES in Europe. These recommendations are based, in part, on a survey of how Member States addressed these concerns in previous HESs or similar studies (Tolonen 2008) and on the experiences from the EHES pilot surveys. In particular, we provide some general recommendations on the ethical conduct of a HES, with specific reference to the safeguarding of privacy (or "data protection"); we also provide and discuss a model of an informed consent form, which is intended as a guide for creating such a form for HESs in Europe. As there are differences in national legislation and guidelines, the form needs to be adapted for national use.

4.1 Legislation and guidelines

National HESs must be conducted according to ethical standards, which, for all research on humans, are regulated by national legislation and national and international guidelines. Key documents at the time of seeking ethical approval for the national surveys and when preparing the national survey protocols are:

1. national acts regulating the status and/or rights of patients;
2. national medical research acts;
3. other national ethical principles of research involving humans;
4. international biomedical research guidelines, such as;

- the Declaration of Helsinki, "Ethical Principles for Medical Research Involving Human Subjects", which is considered to be the pillar of ethical standards (WMA 2008);
- the Belmont Report in 1979, "Ethical Principles and Guidelines for the Protection of Human Subjects of Research" (NIH 1979);
- the Recommendation of the Committee of Ministers No. R(90) 3 concerning medical research on human beings (Committee of Ministers 1990);
- the Oviedo Convention on Human Rights and Biomedicine. (Oviedo 1997);
- Council of Europe in 2005: Additional Protocol to the convention on Human Rights and Biomedicine, concerning Biomedical Research (CEO 2005); and
- Council of International Organizations of Medical Sciences and WHO in 2002: International Ethical Guidelines for Biomedical Research Involving Human Subjects (CIOMS and WHO 2002).

4.2 Role of ethics committees

An ethics committee is a body that is responsible for evaluating research proposals from an ethical standpoint. In particular, this committee, which can be local, regional, or national, evaluates the given proposal in terms of its compliance with national legislation and regulations. The evaluation covers not only the performance of the research itself (i.e., that the participant will not be harmed or placed at risk) but also the contents of the informed consent and how it is obtained, the safeguarding of privacy, and the use of data and biological materials, both for the research being conducted and any future purposes.

The approval of the ethics committee is needed not only for full-size national HES but also for pilot studies. It must also be considered that obtaining ethical approval can be a time-consuming process; in some countries or circumstances it may take up to one year. Therefore, the procedures for obtaining approval need to be started as early as possible, during the beginning of the planning phase.

The general steps for obtaining ethical approval are illustrated in Figure 4.1, though the detailed procedures may vary by country. The first step is to identify the appropriate ethics committee and the documentation that this committee requires for applying for approval. In preparing this documentation, it is recommended that experts in ethical issues be consulted. Once the proposal is submitted for approval, the ethics committee may request modifications if the proposal does not fulfill the established criteria. The HES cannot be started before approval is obtained.
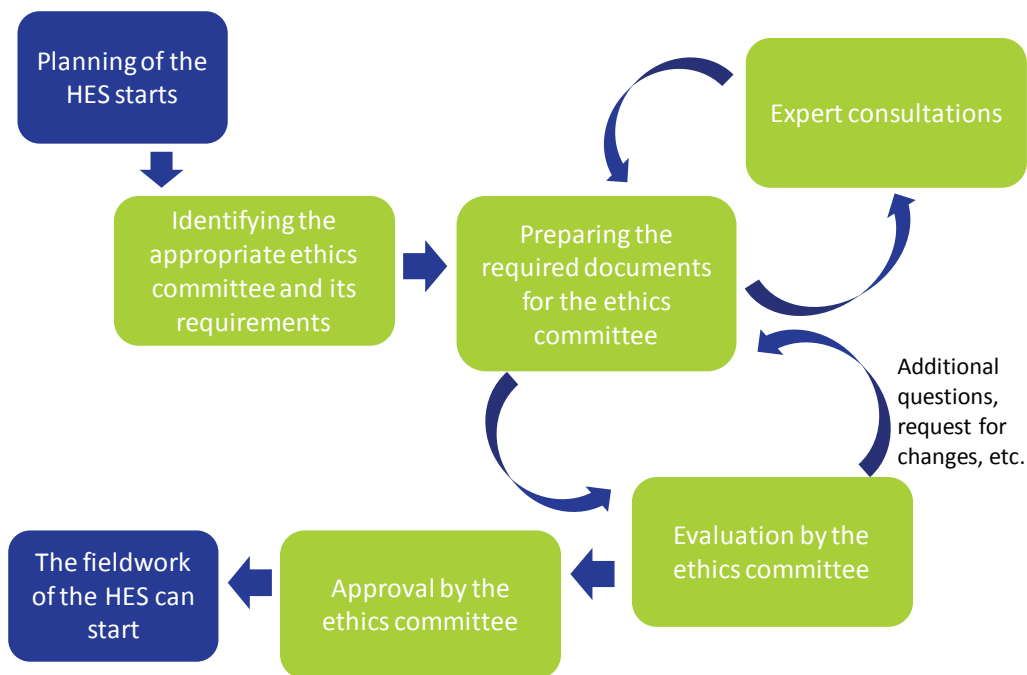
Figure 4.1 Process for obtaining ethical approval

4.3 Data protection

The Declaration of Helsinki states, *"Every precaution should be taken to respect the privacy of the subject and the confidentiality of the patient's information..."*. This issue has become increasingly important in light of the progress made in information technology and the consequent ease of access to data. That privacy is safeguarded is ensured through legislation (generally a "Data Protection Act").

Performing a HES includes collecting individual level data which are also personal data (i.e., sensitive data regarding health). For this reason, the HES protocol must comply with the given country's Data Protection Act and cover all aspects of data protection, in particular: access to data, the exchange of data, record linkage (e.g. linkage to register data on socio-demographic factors and health service use with demographic data available for non-respondents and with survey data for the respondents), and anonymisation procedures (more detailed information on methods for ensuring data security are provided in Part A, Chapter 12).

In safeguarding privacy, the reform of data protection rules in the EU need to be considered. The new EU Data Protection Regulation 2016/679 shall apply from May 2018. The Directive entered into force on May 2016 and EU Member States have to transpose it into their national law by May 2018. The issue of ensuring data protection is also of extreme importance in developing informed consent material.

To understand better the concept of data protection, some commonly used terms are defined below (more detailed definitions are provided in the above-mentioned Directive).

- *Personal Data* - information regarding an identifiable person, that is, one who can be directly or indirectly identified, in particular by reference to an identification number or to factors specific to his/her physical, physiological, mental, economic, cultural or social identity;

- *Processing of Personal Data* - any operation (automatic or not) performed on personal data, for example, collection, storage, adaptation or alteration, retrieval, linkage, destruction and dissemination;

- *Controller* - the person or entity that determines the purposes and means of the processing of personal data;

- *Processor* - the person or entity that processes personal data on behalf of the controller;

- *Personal Data Act (or Data Protection Act)* - legislation for protecting the privacy of natural persons in the processing of personal data;

- *Sensitive Data* - personal data revealing racial or ethnic origin, politicpersonal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation;

- *Right of Access* - the subject has the right to personal data which have been collected concerning him or her, and to exercise that right easily and at reasonable intervals. This includes the right to have access to data concerning their health, for example the data in their medical records containing information such as diagnoses, examination results, assessments by treating physicians and any treatment or interventions provided.

4.4 Informed consent

4.4.1 Objectives of informed consent

Before performing any kind of research involving humans, informed consent must be obtained. The objective of informed consent is to allow a person to make a truly informed decision as to whether or not to participate in the HES. In other words, obtaining informed consent goes beyond getting an individual to sign a form: it is a process of communication between an individual and the HES personnel. Its goal is to ensure that the individual fully understands the scopes of the study, the methods adopted, and how the data will be used.

The first step in obtaining informed consent is to provide the study candidate with information. Given that the ultimate goal is to ensure that participants are truly informed, it is fundamental that this information be complete and clear. This is also important in terms of the HES participation rate, in that unclear or poorly written material (or

even material that relies too heavily on "scientific" terms, which could be intimidating) could result in an invitee's decision not to participate.

4.4.2 Means of providing information for informed consent

In this section, we discuss how to provide the HES invitee with information on the HES and how to obtain written informed consent. These procedures basically consist of three main activities: i) making the person aware of the HES; ii) providing the person with a clear understanding of what participation involves; and iii) obtaining the person's signature attesting to his/her consent to participate. This is done using what is called an "informed consent form". This form contains all of the information required by the HES invitee for understanding the HES and what participation entails and a space for the personal's signature, attesting to the fact that he/she has understood the information and agrees to participate. In some cases, the information on the HES is contained on a separate document (referred to as an "information notice" or "information leaflet"). Furthermore, an "invitation letter" can also be provided, which is used as an introduction to explain in general what the study is about, its importance, and how and when the invitee will be contacted (If used, the invitation letter should be brief yet "appealing"; for more information on the invitation letter, see Part A, Chapter 13.).

One advantage of using separate documents is that the information notice and invitation letter can be provided some time before the HES invitee provides a signature, so that he/she has sufficient time for reading and understanding the information before agreeing to participate. The choice of the information material's format also depends on such factors as the general organization of the study and the national legislation and regulations regarding privacy or related issues. For example, in some countries, the ethics committee explicitly requires that the informed consent form consist of a single document that includes both the necessary information and the signature. In developing informed consent material, setting up telephone help-lines for answering invitees' questions and providing clarifications can also be considered, as can the translation of material into other languages. A web-site dedicated to the HES could also be created, with all of the information about the study, including the information notice itself.

EXAMPLES OF INFORMED CONSENT MATERIAL AND WHEN TO PROVIDE IT

Given below are three examples of formats for the informed consent material and when to provide it. The extent of the information in each document can be modified, depending on specific needs. For example, if the information notice is "extensive" (i.e., if it contains most of the necessary information), then the invitation letter and the informed consent form can be brief. Whatever format is chosen, the documents should be complementary, that is, there should not be excessive overlap, to avoid burdening the candidate with too much reading.

Example 1

Invitation letter + extensive information notice, sent together some days (or weeks) before the invitee has an appointment to the HES
⇨ brief informed consent form, provided at the beginning of the examination visit (before any measurements)

Example 2

Invitation letter + brief information notice, both sent some days (or weeks) before the invitee has an appointment to the HES
⇨ extensive informed consent form, provided at the beginning of the examination visit (before any measurements)

Example 3

Invitation letter sent some days (or weeks) before the invitee has an appointment to the HES + extensive information notice published on a web-site, with toll-free line provided to ask questions
⇨ written information notice (the same published on the web-site) and informed consent form provided and explained to the candidate at the beginning of the examination visit (before any measurements)

Example 4

Invitation letter + extensive information notice + full informed consent form, sent together some days before the invitee has an appointment to the HES. The person is asked to bring the informed consent form to the examination site.

4.4.3 Recommendations for creating an informed consent form

This section is intended to help the survey organizers to create an informed consent form for the national HES. This should be adapted for national needs. For example, in the present form, invitees are asked to provide a single signature which indicates consent to participate in all parts of the study. However, in some countries it may be required (or preferable) that the participant provides a separate consent and signature for each individual activity or some specific parts of the survey (e.g., blood taking, linking of data to other databases). Many of the statements on this form are followed by a comment (in italics) that provides suggestions or considerations which may help to adapt the form for use in the national HES.

Given that the ultimate goal is to ensure that participants are truly informed, the information provided must be complete and clear. Terminology that is simple and easy to understand should be used, avoiding scientific terms when possible. Moreover, excessively complex or long descriptions can confuse or intimidate study invitees.

The protocol for conducting the HES in each country, including the informed consent form, will have to be approved by the national, regional, or local ethics committee, so as to ensure that it complies with national legislation and ethical standards. Many of the sections in this form may have to be modified to be consistent with the legislation in each country (e.g., regarding access to data and storage of samples in biobanks).

4.4.4 Model of an informed consent form to be used in European HESs

The model provided below includes an introduction which explains its purpose and provides recommendations for those who will be responsible for this aspect of the HES. In the model, the information notice and the signature form are a single document, yet as mentioned above, these can be two separate documents provided to study invitees at different times, together with an invitation letter.

In the model, asterisks indicate items that are "mandatory", that is, those that should always be included. The other items depend on the specific characteristics of the HES. For example, if data linkage is not performed, then a specific consent is not needed.

See Appendix 4.1 for Template of an informed consent form

a) Introductory information on the HES

To the survey invitee,

You are invited to take part in a National Health Examination Survey (or "HES").

This study is carried out for obtaining information on general health by interviewing individuals and by measurements that can be important to health. This information is used to acquire knowledge on the health status of the population, which will be important in promoting and improving the health of all and to develop health services.

> **Comment:** *Information on health concerns that are important in the specific country and for which a HES could be beneficial can be added here. For example "In Italy, obesity is becoming an increasingly important health concern, yet there is little information on what percentage of the population can be considered as obese." If the invitee feels that the study would be socially useful, then the chances of him/her participating could increase.*

The HES is being conducted by (specify name of organization conducting the HES in your country) among a sample of (specify expected number of participants individuals in specify study area, such as the town or province). Your name was chosen from (specify source of the person's name and the area to which it refers)

Comment: This sentence should specify how the individual was chosen (e.g., from electoral rolls, social insurance registers, population registers), so that he/she is aware of how the research personnel obtained his/her name.

All aspects of this study have been approved by the Ethics Committee of the (specify the name or level of the ethics committee). The present form includes important information about the study and a description of what will be asked of you if you decide to participate. In order to participate, you will need to carefully read and sign this form. If any part of this form is not clear to you, please feel free to ask the person/s receiving the informed consent.

Comment: The wording of this sentence may change according to who is available for providing clarifications or depending on whether or not information aids, such as telephone help-lines, are provided.

Your participation is important to us, but please be assured that it is voluntary, that you may leave the study at any time, and that your data will be kept confidential.

b) Collection of personal data

During the survey, you will be asked to answer questions on ...

Comment: Specify the topics that the questions will cover. If an interview is not conducted (e.g., if a self-administered questionnaire is used), the wording of this section should be modified accordingly).

Measurements of your height, weight, waist circumference and blood pressure will be taken; blood/urine/saliva samples will also be taken.

Comment: If the HES comprises additional modules, then modify accordingly.

These samples will be tested for ...

Comment: To be modified in accordance with the specific objectives of the HES.

Comment: It may be important to assure study invitees that the samples will not be used to test for other purposes (e.g., HIV testing, drug testing) which they would not approve; examples could be provided. If DNA testing is performed or additional tests are anticipated to be made later for additional scientific purposes (e.g. biobank use), this should be explicitly declared.

To perform the interview and the physical examinations and to collect the samples needed for the survey, approximately ___hours of your time will be needed. These activities will be performed in __visits.

Comment: Specify the total time in hours, number of visits, and the amount of time per visit. This is an important consideration

for invitees in deciding whether or not to participate. The time needed should not be underestimated.

c) Information on risks

Comment: Given that the risk associated with the taking of blood samples is minimal, this section can be eliminated, although in certain countries it may be necessary to make such a statement. However, if any activities that may pose a risk are added to the HES, the potential risks must be disclosed.

Comment: To reassure the invitee, the following sentence may be included: "All examinations are conducted by qualified and specially trained nurses/doctors; they are also trained to react competently to unforeseen situations".

Comment: If insurance coverage is provided for the duration of the stay of the participant at the study centre, then this should be stated.

d) Compensation

For your participation in this survey, you will receive....

Comment: If no compensation is to be provided, then it is possible to write "You will not be paid for taking part in this study." or to eliminate this statement. If instead it is provided, the description of compensation must be clear. Payment or other forms of incentive may not be allowed in certain countries. It may also be useful to specify that all tests are carried out free of charge for the participant.

e) Use of results

Your personal results will be reported to you in approximately ____ weeks/ _____ months.

Comment: If the participant's general practitioner is responsible for providing the results to the participant, then this should be specified and the name and address of the GP should be checked with the participant If sensitive and potentially upsetting tests are planned (e.g. HIV tests) or additional tests will be carried out later from the stored samples, this raises an additional ethical issue. The procedure of reporting back such tests needs to be planned carefully in each country, taking also into account the possibility that there may be a laboratory error.

The data collected from you will be stored and used for research purposes by the (specify name of institution conducting the HES). Only limited number of data managers will have access to the full database.

The anonymized data will be used by researchers and will also be provided to researchers in other institutes collaborating on the survey, possibly in other countries. The anonymized data can also be combined

with data from the HES conducted in other European countries in a centralized database.

> **Comment:** *Given that data protection laws may vary by country, the institutes with access to data may differ. For example, in some countries it may be legal to provide data on individuals to general practitioners. It is important that the participant is aware of who will have access to his/her data.*

f) Record Linkage

The data may also be combined (or "linked") with other data from different sources to study the relationships of specific diseases with the risk factors determined in this survey; such studies are important for improving the prevention of the disease.

> **Comment:** *The databases (data sources), e.g. the national hospitalization register, if known, should be specified.*

g) Confidentiality/Privacy

The data collected will be kept strictly confidential. They will be stored, analysed and handled in accordance with legislation on Data Protection and Privacy. No information that could be used to identify you will be provided to third parties. The results of this study will be published , but the publications will not include any information that could lead to participants' identification.

> **Comment:** *Describe procedures that will be followed to keep subject information and specimens secure and confidential. For example: "To ensure that the data collected from you remain confidential and that your privacy is protected, records will be kept in a separate research file that does not include names or other information that could be used to identify you. Specific national legislation or regulations on Data Protection and Privacy could also be provided here.*

If you withdraw from the study, you may decide that your data and the samples will not be used / will be eliminated.

> **Comment:** *Whether or not data from persons withdrawing from the study must be discarded depends on the specific legislation in the given country.*

The person/entity responsible for safeguarding privacy in this study is [specify]

Permission to perform this study has been provided by [specify Data Protection Authority].

> **Comment:** *Depending on the Data Protection Act, it might be necessary to notify or request permission from the Data Protection Authority.*

At any point during or after the study, if you are concerned about a possible violation of your privacy or about any other issues regarding your data, you can contact *[specify name and contact information of the person/entity responsible for privacy in this study]*.

h) Long-term storage

Your samples may be stored at the (*Specify name of organization conducting the HES*) or in what is referred to as a "biobank" (that is, a long-term storage facility for biological materials) and used at a later time for other health studies.

> **Comment:** *In the given country, there may be legal limitations regarding the storage (including duration) and use of biological materials. Keep in mind that the term "biobank" may be intimidating for some and that terms such as "long-term storage" may be more suitable.*

i) Additional studies/Follow-up/Specific parts of the study

> **Comment:** *If needed, specific consent for the following should be asked.*

After this survey is complete, we may want to re-contact you for more questions and other examinations; Do you agree to be re-contacted (please note that this could even be in a few years)?
☐ Yes
☐ No

Do you consent to record linkage (*specify the data sources*)
☐ Yes
☐ No

Comment: *The databases that are to be linked, if known, should be specified.*

Do you consent to the long-term storage of your samples/your samples stored in the biobank (*specify name*)?
☐ Yes
☐ No

j) CONTACT INFORMATION ABOUT THE PRESENT STUDY

For any questions or concerns, you can contact the researcher(s) listed below.

> **Comment:** *It is important that the participant be provided with the possibility to speak with someone for any questions or doubts that he/she may have. Not only can this be reassuring for the study candidate or participant, but it might also increase the participation rate.*

Principal Investigator: specify name of Principal Investigator

> **Comment:** *The person available for providing clarifications may change according to how the HES is organized.*

E-mail:
Mailing Address:
Telephone:

Consent

Participant:

I understand the information printed on this form. I understand that if I have more questions or concerns about the survey or my participation, I may contact the person(s) listed above.

> **Comment:** *This section can be modified to emphasise the interactive aspects of informed consent, for example: "I have read and understood all of the information regarding this study, which has also been verbally explained, and all of my questions have been adequately answered."*

Signature of participant: _____
Date: _____
Name (Print legal name): _____
Participant ID: _____

Legal Representative (if applicable):

> **Comment:** *If persons unable to fully consent for themselves are included in the HES, this section should be filled in by the person's legal guardians.*

Signature of person legally authorized to give consent

_____
Date: _____
Name (Print name): _____

Relationship to participant:
☐   Parent
☐   Spouse
☐   Son/Daughter
☐   Sibling
☐   Legal Guardian
☐   Other: _____

Reason participant is unable to sign for himself/herself:
_____

Person receiving the informed consent:

I have received the informed consent of (name of participant).

> **Comment:** *A sentence can be added to emphasise the interactive aspects of informed consent, for example:"I have informed the participant of the objectives and conduct of this study and*

of its compliance with data protection procedures, both verbally and in writing."

Signature of person receiving informed consent:
_____
Date: _____
Name (Print legal name): _____

References

- CIOMS & WHO 2002: International Ethical Guidelines for Biomedical Research Involving Human Subjects. Available from http://www. cioms.ch/index.php/publications/printablev3/541/view_bl/65/bioethics-and-health-policy-guidelines-and-other-normative-documents/23/ international-ethical-guidelines-for-biomedical-research-involving-human-subjects?tab=getmybooksTab&is_show_data=1; last accessed 7 November 2016.

- Committee of Ministers (1990) Recommendation No. R (90) 3 of the Committee of ministers to member states concerning medical research on human beings. Available at https://wcd.coe.int/wcd/com.instranet. InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=56 9941&SecMode=1&DocId=590274&Usage=2. Accessed on 6 June 2011

- Council of Europe (2005) Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research. Available from http://www.coe.int/en/web/conventions/full-list/-/conventions/ rms/090000168008371a; last accessed 7 November 2016

- Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union 2016;59:1-88.

- NIH (1979) The Belmont Report. Ethical Principles and Guidelines for the protection of human subjects of research. Available at https://videocast. nih.gov/pdf/ohrp_appendix_belmont_report_vol_2.pdf; last accessed 7 November 2016;

- Oviedo (1997) the Oviedo Convention on Human Rights and Biomedicine. Available from https://rm.coe.int/CoERMPublicCommonSearchServices/Di splayDCTMContent?documentId=090000168007cf98; last accessed 7 November 2016

- Tolonen H, Koponen P, Aromaa A, Conti S, et al (eds.) Recommendations for the Health Examination Surveys in Europe. Publications of the National Institute for Public Health, Helsinki 2008. B21/2001. Available at http:// urn.fi/URN:NBN:fi-fe201204193358. Accessed on 22 January 2013.

- WCD. the Recommendation of the Committee of Ministers No. R(90) 3 concerning medical research on human beings (WCD) Available from https:// wcd.coe.int/wcd/com.instranet.InstraServlet?command=com.instranet. CmdBlobGet&InstranetImage=569941&SecMode=1&DocId=590274&Usa ge=2; last accessed 7 November 2016

- WMA (2008) Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. Available from http://www.wma.net/ en/30publications/10policies/b3/; last accessed 7 November 2016

Appendix 4.1 Template of an informed consent form

National Health Examination Survey (NHES)

To the survey invitee,

You are invited to take part in a National Health Examination Survey (or "HES").

This study is carried out for obtaining information on general health by interviewing individuals and by measurements that can be important to health. This information is used to acquire knowledge on the health status of the population, which will be important in promoting and improving the health of all and to develop health services.

The HES is being conducted by (specify name of organization conducting the HES in your country) among a sample of (specify expected number of participants individuals in specify study area, such as the town or province). Your name was chosen from (specify source of the person's name and the area to which it refers)

All aspects of this study have been approved by the Ethics Committee of the (specify the name or level of the ethics committee). The present form includes important information about the study and a description of what will be asked of you if you decide to participate. In order to participate, you will need to carefully read and sign this form. If any part of this form is not clear to you, please feel free to ask the person/s receiving the informed consent.

Your participation is important to us, but please be assured that it is voluntary, that you may leave the study at any time, and that your data will be kept confidential.

During the survey, you will be asked to answer questions on ...

Measurements of your height, weight, waist circumference and blood pressure will be taken; blood/urine/saliva samples will also be taken.

These samples will be tested for ...

To perform the interview and the physical examinations and to collect the samples needed for the survey, approximately ____hours of your time will be needed. These activities will be performed in __visits.

For your participation in this survey, you will receive....

Your personal results will be reported to you in approximately ____ weeks/ _____ months.

The data collected from you will be stored and used for research purposes by the (specify name of institution conducting the HES). Only limited number of data managers will have access to the full database.

The anonymized data will be used by researchers and will also be provided to researchers in other institutes collaborating on the survey, possibly in other countries. The anonymized data can also be combined with data from the HES conducted in other European countries in a centralized database.

The data may also be combined (or "linked") with other data from different sources to study the relationships of specific diseases with the risk factors determined in this survey; such studies are important for improving the prevention of the disease.

The data collected will be kept strictly confidential. They will be stored, analysed and handled in accordance with legislation on Data Protection and Privacy. No information that could be used to identify you will be provided to third parties. The results of this study will be published , but the publications will not include any information that could lead to participants' identification.

If you withdraw from the study, you may decide that your data and the samples will not be used / will be eliminated.

The person/entity responsible for safeguarding privacy in this study is [specify]

Permission to perform this study has been provided by [specify Data Protection Authority].

At any point during or after the study, if you are concerned about a possible violation of your privacy or about any other issues regarding your data, you can contact *[specify name and contact information of the person/entity responsible for privacy in this study].*

Your samples may be stored at the (*Specify name of organization conducting the HES*) or in what is referred to as a "biobank" (that is, a long-term storage facility for biological materials) and used at a later time for other health studies.

After this survey is complete, we may want to re-contact you for more questions and other examinations; Do you agree to be re-contacted (please note that this could even be in a few years)?
☐ Yes
☐ No

Do you consent to record linkage (*specify the data sources*)
☐ Yes
☐ No

Do you consent to the long-term storage of your samples/your samples stored in the biobank (*specify name*)?
☐ Yes
☐ No

For any questions or concerns, you can contact the researcher(s) listed below.
Principal Investigator: specify name of Principal Investigator

E-mail:
Mailing Address:
Telephone:

Consent

Participant:

I understand the information printed on this form. I understand that if I
have more questions or concerns about the survey or my participation,
I may contact the person(s) listed above.

Signature of participant: _____
Date: _____
Name (Print legal name): _____
Participant ID: _____

Legal Representative (if applicable):

Signature of person legally authorized to give consent
_____
Date: _____
Name (Print name): _____

Relationship to participant:
☐ Parent
☐ Spouse
☐ Son/Daughter
☐ Sibling
☐ Legal Guardian
☐ Other: _____

Reason participant is unable to sign for himself/herself:
_____

Person receiving the informed consent:

I have received the informed consent of (name of participant).

Signature of person receiving informed consent:
_____
Date: _____
Name (Print legal name): _____

5. Selecting the questionnaire modules, measurements and biological samples

PART A

Johanna Mäki-Opas[1], Kari Kuulasmaa[1], Päivikki Koponen[1], Laura Paalanen[1], Hanna Tolonen[1]

[1] National Institute for Health and Welfare (THL), Helsinki, Finland
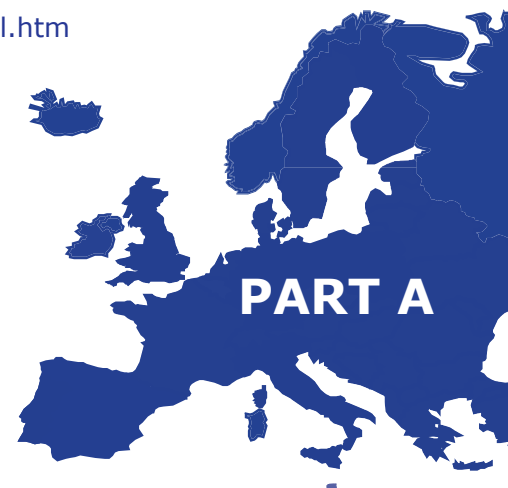
EHES collects data through questionnaires, physical measurements, and analysis of biological samples. This chapter outlines those components as well as the importance and rationale of them. Measurements have been divided into core and additional measurements. Core measurements are a minimum set of measurements which should be included in every national HES. When the country has more experience, funding and national need for information, additional measurements can be added. There is a list of measurements included in the previous national HESs on the EHES web site (http://www.ehes.info/national/national_hes_measurements.htm)

5.1 Selection criteria

The selected questionnaire modules, measurements and biological samples should be based on objectives of the survey and specified research questions as well as the analysis plan (de Bruin 1996). It is important to review all measurements carefully to make sure that they are really needed and that they provide required valid information for the selected indicators. See list of core indicators recommended for the EHES in Part C, Section 4. Indicators. One measurement may contribute to several indicators (Tolonen 2005).

Table 5.1 provides the criteria which have been used for the selection of the EHES core measurements. Also the additional measurements should be evaluated against these criteria. For the additional measurements it is recommended to have at least one measurement which people are interested in and which motivates people to take part in the survey. This may increase the participation rate. It is also useful to check that all personal results can be interpreted for the participant and used to estimate needs for care and preventive activities for the individuals as well as for the population.

Table 5.1 Criteria for selecting the measurements for a national HES (modified from Primatesta et al. 2008, Tolonen 2005)

Criteria	Rationale
Public health importance	The measurements should address key public health problems and be common enough to be measured in the population. Rare phenomenon should not be included in the survey as only few cases can be identified.
	The measurements should support monitoring needs defined in the international agreements such as the WHO Action Plan of the European Strategy (WHO 2012) and the WHO Global Action Plan (WHO 2013) for the Prevention and Control of Noncommunicable Diseases and the European Core Health Indicators (ECHI).
	At the national level the following can also be considered:
	• Possibilities to evaluate health and economic implications of national policies and action plans
	• Possibility to demonstrate economic effect through association of the measured phenomenon to work absences, early retirement, as well as need and use of health care services.
Clear interpretation of the results	The measurements need to provide reliable information about the phenomenon which can be given as feedback to the survey participants and which can be used in health monitoring, health service development and health policy evaluation.
Availability of international standards	Internationally standardized measurement protocols/questionnaire modules should be used whenever possible to ensure validity and to enhance comparability of the results between countries and over time (at least if other issues jeopardizing comparability, such as differences in data collection modes can also be minimized). For questionnaire modules, use of internationally standardized instruments ensures that the questions have been validated. However their national validation (validation of the translation and evaluation of feasibility) is always needed.
Practicality, easy to administer	The measurements should be feasible to conduct in relation to available time, equipment and its transportation and calibration, as well as training and qualifications of the personnel.
Surveys as the primary source of information	Information about the phenomenon cannot be obtained as reliably through other existing data sources (such as administrative registers).

Criteria	Rationale
Cost of the survey	Costs of measurements and available funds need to be in balance. Selecting one expensive measurement may drop out several cheaper ones.
Ethical acceptability	Measurements have to be ethically approved and safe for the participants, as well as accepted by health care professionals. If deviations from normal values are identified, access to care and preventive activities needs to be assured.
Acceptability to the participants	The selected measurements should not be too time consuming, causing extra burden, pain or discomfort for the participants.

5.2 Measurements

The core EHES questionnaire items, physical measurements, and analyses of blood samples collect data mainly on major chronic diseases (mainly cardiovascular diseases and diabetes), and their risk factors (e.g. obesity, high blood pressure and high serum cholesterol). These diseases and risk factors are preventable at both individual and community level (Vartiainen 2009).

The WHO global action plan for the prevention and control of non-communicable diseases 2013-2020 has defined sevaral indicators which should be monitored and for which data can be obtained reliably only through health examinaition survey. These indicators are based on questions on smoking, alcohol use, and physical activity; physical measurements of blood pressure, and height and weight, and analysis of biological samples for blood glucose, total cholesterol and sodium intake. (WHO 2013)

5.2.1 The EHES core questions

The EHES core questionnaire items which are based on the EHIS questionnaire should be administered as recommended in EHIS. The following documents should be reviewed while preparing the national version of the EHES questionnaire and when training the fieldwork personnel:

- European Health Interview Survey (EHIS wave 2) Methodological manual includes questions, interviewer guidelines and other related instructions. (Eurostat 2013)
- Rough translations of the questions can be found from the translated versions of the Commission Regulation No 141/2013 (EU 2013)

Guidelines and quality criteria for EHIS questionnaire administration have also been documented (Davidsson et al 2009, EHIS wave 2 Methodological Manual 2013, see also the EHIS validation rules).

The EHES core questions are mostly questions that are necessary for the reporting and interpretation of the data from the physical measurements and biological samples. Whenever possible, EHIS questions are used. In the EHES core questionnaire some questions have been slightly modified from EHIS and there are also a few other than EHIS questions. Two waves of EHIS have been conducted by 2016 and planning of the wave 3 is ongoing. There may be some changes to the existing questions in wave 3. This revision is expected to be completed by 2018. The EHES core questionnaire includes questions on:

- Household size
- Sex
- Age
- Marital status
- Socio-economic status
  - Education
  - Occupation
  - Household income
- Self-reported height and weight
- General health
- Chronic diseases
- Use of medication
- Smoking

All the questions can be found in Part B, Section 5.7.

Age and sex enable reporting of the HES results by sex and age group and the age-adjustment of the results for comparison between populations. Education, occupation and household income are needed for the estimation of socio-economic differences in the population. Some countries may have the possibility to obtain these demographic and socio-economic data through the sampling frame or linkage with registry data in which case there is no need to ask them. However, if such data linkage is not possible in the country or if the coverage of the registry data is incomplete, it is important to include these questions into the survey. If feasible, questions on self-reported ethnicity and/or country of origin should also be asked.

Even though height and weight will be measured, they should also be asked. All information obtained from the selected persons is important especially when statistical adjustment for the non-participation is done (see Part A, Chapter 13.). This enables for example the analysis of non-participant's BMI in a case that participant fills in only the questionnaire but does not take part to the physical examinations. Asking height and weight also enables to estimate the differences of measured and self-reported height and weight between countries, and by sex, age and socio-economic status, if the questionnaire is filled in before the measurements.

The three questions forming the Minimum European Health Module (MEHM) which is expected to be included in all European social and health surveys are also included in the EHES questionnaire. The structural indicator Healthy Life Years is calculated on the basis of the questions of MEHM. More generally the three questions are used for the calculation of the prevalence of perceived health, self-reported long-standing illnesses or health problems and long-term activity limitations (the Global Activity Limitation Indicator, GALI). The GALI question has been under review in the project on standardisation of social variables, and may be modified for future EU surveys. The questions on chronic diseases measure the main public health concerns, which are also a major reason for using the health care services. Measuring chronic morbidity is useful for overall evaluation of health status in the population. It is also useful for evaluation, policy formulation and assessment of needs for health care. Register based data on chronic morbidity often suffers from low comparability and under coverage (due to not all patients using the services or receiving the benefits covered in the register). The answers to the questions on specific diseases are needed from the same person as for whom the physical measurements are performed. For example the question on the use of hypertension medication is combined with the results of blood pressure measurement to see how well hypertension is treated and controlled in the population. Also the cholesterol and glucose levels in blood are combined with the related questions in the core indicators recommended for the EHES.

Smoking is an important factor for lung diseases and cancer, other cancers and diseases of the circulatory system. Lung, trachea and larynx cancer is the type of cancer with the higher standardized death rate among men in EU. In addition, important policy activities are developed at EU level in order to limit tobacco consumption and many of the Member States have forbidden or are in the process of forbidding smoking in working places and public areas. Smoking is a major determinant of chronic diseases and other risk factors covered in the EHES core measurements, as well as of other health outcomes.

5.2.2 The core physical measurements

Physical measurements are needed because self-reported data is often not sufficient to assess population levels and trends or to make comparisons between populations. Self-reported information is prone to recall and awareness bias. (Tolonen et al 2014) These selected measurements have also been widely measured in previous national HESs conducted in Europe (Tolonen 2008).

The core physical measurements are:

- Height
- Weight
- Waist circumference
- Blood pressure

Body Mass Index (BMI) is a widely used indicator of obesity. It is defined as body weight divided by the square of height. According to the Global Burden of Disease Study 2015, the prevalence of overweight and obesity has steadily increased (GBD 2015 Risk Factors Collaborators, 2016). The increase in obesity and overweight among the population is one of the most important public health issues in developed countries. Overweight and obesity are a major risk factors for diseases of the circulatory system, diabetes and several other chronic diseases (Malnick 2006, Bastien et al 2014). The evolution of the way of life and food consumption in the EU Member States is characterized by low physical activity and energy-dense food intake which increase the body mass index.

Waist circumference is used as an indicator of abdominal obesity. Since increasing evidence has shown that waist circumference reflects the accumulation of visceral fat better than waist-to-hip ratio, the waist circumference is the preferred measure in population studies (Seidell 2001). Waist circumference is significantly associated with the risk of incident CVD events and type 2 diabetes (de Koning 2007, WHO 2011a).

Measuring blood pressure gives the prevalence of actual and potential hypertension. In 2015, hypertension was the leading global risk factor for disability-adjusted life-years (DALYs) (GBD 2015 Risk Factors Collaborators, 2016). Single-occasion blood pressure measurement has been shown to be a strong indicator of coronary and cerebrovascular risk (MacMahon 1990). However, the diagnosis of hypertension requires follow-up and observed high blood pressure on several occasions which is not feasible in a national HES.

5.2.3 The core biological samples

The EHES surveys include the collection and analysis of biological samples. The core blood samples are:

- Non-fasting blood samples
  - Total cholesterol
  - HDL cholesterol
  - Glycated haemoglovin (HbA$_{1c}$)
- Fasting blood sample (8-14 hours)
  - Glucose

High serum total and HDL cholesterol are major risk factors of cardiovascular diseases. Increased glucose level or HbA$_{1c}$ may indicate insulin deficiency or insulin resistance which indicates risk for diabetes. (Emerging Risk Factors Collaboration 2009.)

Because of potential difficulties in requiring fasting from all participants, the glucose measurement may cover only a sub-sample of the survey. It should be noticed that the fasting should last at least four hours, preferably eight, but not more than 14 hours. Alternatively, HbA$_{1c}$ could be measured from non-fasting samples (Sherwani 2016). When

HbA1c is used, fasting blood glucose should ideally also be measured in a subsample of participants to provide information about how the two tests relate (NCD Risk Factor Collaboration 2015).

The classification and diagnosis of type 2 diabetes has relied on the measurement of fasting plasma glucose concentrations or oral glucose tolerance tests. Interpretation of non-fasting glucose values is difficult, if not impossible for classification of diabetes in large population studies. On the other hand, obtaining blood samples from adequately fasted participants is often impractical in health surveys.

Glycated haemoglobin, HbA_{1c}, reflects the time-averaged blood glucose concentration during the previous 2-3 months. There is a close relationship between HbA_{1c} and glucose. Therefore, it has been proposed to substitute plasma glucose with HbA_{1c} not only for following the effectiveness of diabetes treatment but also for classification of type 2 diabetes (The International Expert Committee 2009). Its superiority over plasma glucose, especially in health surveys, lies in that its measurement does not require a fasting blood sample.

In the past, measurement of HbA_{1c} has been hampered by the measurements not having been standardized to a sufficient level. Recently, however, a consensus statement on the worldwide standardization of the HbA_{1c} measurement has been published (Hanas 2010). It is foreseen that in the very near future HbA_{1c} could replace plasma glucose as a core measurement. Therefore, measurement of HbA_{1c} is strongly recommended already now. However, the higher cost of HbA_{1c} than plasma glucose analysis is still a problem (WHO 2011b, Bonora 2011).

5.3 Additional measurements

In addition to the core measurements, countries may include additional questionnaire modules, physical measurements, and collection of biological samples into the national HES. When choosing the additional measurements, the criteria shown in Table 5.1 should be followed. Countries with little experience from earlier HESs are recommended to keep the number of additional measurements low to allow adequate planning and preparation for all measurements and fieldwork procedures. Experienced countries may include a wide range of additional measurements to the survey if they are confident that they can manage the survey process and they have sufficient funding. Additional measurements can be added to the survey as modules that are relevant for example to specific sub-groups of the population, such as certain age groups, ethnic groups or other sub-populations of regional/local interest.

When selecting additional measurements, the countries should consider their implications to the survey administration, the time taken for training, to administration of the questionnaire and carrying out the physical measurements as well as the costs and the periodicity of the survey. If the survey will be repeated frequently, different additional modules can be considered for each round of data collection. The annual Health Survey for England focuses on different single or multiple health issues and/or population subgroups in different years, and the

samples may be boosted to make the study of specific subgroups of the population possible (Mindell et al 2013). In a national survey with large samples, an additional measurement module for a smaller sub-sample may be included, like in the German DEGS survey (Scheidt-Nave et al 2012) and the previous Finnish FINRISK and Findiet surveys (Reinivuo et al 2010). When the survey will be carried out less frequently, it may be feasible to build a more comprehensive survey covering several health topics. A commonly used target is that the physical measurements should be limited to take one hour. Some evidence suggests that longer surveys are less acceptable to respondents. But there are also experiences (e.g. the Health 2000 and Health 2011 surveys in Finland) where a more comprehensive survey with long examinations has been attractive to the participants as it gives more information on their own health (Lundqvist & Mäki-Opas 2016).

5.3.1 The additional questions

Additional questions/questionnaire modules may be needed related to the additional physical measurements or biological samples. They may also cover information needed to meet the survey aims and purposes. EHIS questions are recommended to be used when suitable questions are available.

There are many internationally standardized questionnaire modules for example for physical activity, mental health and alcohol consumption.

5.3.2 Additional physical measurements

Measurement protocols for following additional measurements have been prepared and can be found in Part B, Chapter 5 of the EHES Manual:

- Hip circumference
- Handcrip test
- Chair stand test

For other additional measurements, EHES recommendations are not currently available. The countries planning to include measurement for which there is no standard available in the EHES Manual are encouraged to be in contact with survey organizers from other countries to see if they already have a protocol for those specific measurements or are planning to include them to their future surveys. This way the countries interested in the same measurements can collaborate on preparation of the protocols. The EHES RC should also be informed . In this way, also unintentional use of different procedures in countries can be avoided.

In the EHES web site (http://www.ehes.info/national/national_hes_measurements.htm) there is a list of measurements included in the previous national HESs which has also potential additional physical measurements.

5.3.3 Additional biological samples

It is recommended that countries collect more blood samples than are needed for the core analyses. Once suitable blood samples have been collected in the survey and stored properly, they can be used for various measurements in the future (e.g. via national biobanks), if ethical approval and participants' consents for the storage and future analysis are obtained.

From additional blood samples, following issues can be considered:

- Many countries may want to assess serum triglycerides, which are an indicator of cardiovascular risk. Furthermore, triglycerides, together with total and HDL cholesterol can be used to estimate LDL cholesterol, a major risk factor for coronary heart disease. The measurement of triglycerides is complicated by the fact that fasting will be required before blood sampling.

- The measurement of apolipoproteins A1 and B are under consideration for core measurements. They are correlated with HDL cholesterol and LDL cholesterol respectively, and there are indications that they predict cardiovascular diseases better that HDL and LDL cholesterol. (Florvall 2006, Sierra-Johnson 2009) These measurements are easier to standardize than HDL cholesterol and much easier than LDL cholesterol. Furthermore, fasting is not required.

- Countries may also want to collect samples of whole blood for DNA. This will increase the future research potential of the survey, as today the poor availability of large population studies with DNA is a major limitation of genetic research. The DNA collection will imply additional ethical requirements for the survey.

- Many other measurements, such as nutritional biomarkers, environmental exposures, antibodies for infectios diseases and possible new emerging measurements can be done from the stored samples.

Additional to blood samples, urine (spot and/or 24 hours), saliva, and hair/nail samples could be collected. A protocol for the collection of spot and 24 hours urine samples have been prepared and can be found in Part B, Chapter 5 of the EHES Manual.

References

- Bastien M, Poirier P, Lemieux I, Després J-P. Overview of Epidemiology and Contribution of Obesity to Cardiovascular Disease. Prog Cardiovasc Dis 2014;56:369-381.

- Bonora E, Tuomilehto J. The pros and cons of diagnosing diabetes with A1C. Diabetes Care. 2011;34 Suppl 2:S184-90.

- de Bruin A, Picavet HSJ, Nossikov a (Eds.) Health interview surveys: Towards international harmonization of methods and instruments. WHO Regional Publications, European Series, No. 58

- ECHI: http://ec.europa.eu/health/indicators/echi/list/index_en.htm

- EU. Commission Regulation (EU) No 141/2013 of 19 February 2013 implementing Regulation (EC) No 1338/2008 of the European Parliament and of the Council on Community statistics on public health and health and safety at work, as regards statistics based on the European Health Internview Survey (EHIS).

- European Health Interview Survey (EHIS). 1st round 2007-2008. Background and rationale of the questions. European Commission. Available at https://circabc.europa.eu/sd/d/150a71c0-7a17-407d-8d20-2fdc284b0bf7/EHIS%20-%201st%20round%202007-2008%20-%20Background%20and%20rationale%20of%20the%20questions.pdf

- European Health Interview Survey (EHIS). 1st round 2007-2008. Conceptual translation cards and guidelines. European Commission. Availableat https://circabc.europa.eu/sd/d/671d5958-7928-471c-a7b9-0194c34b9cb8/EHIS%20conceptual%20card%20and%20guidelines.pdf

- Eurostat. European Health Interview Survey (EHIS wave 2). Methodological manual. Luxembourg: Publications Office of the European Union, 2013. Available from: http://ec.europa.eu/eurostat/documents/3859598/5926729/KS-RA-13-018-EN.PDF/26c7ea80-01d8-420e-bdc6-e9d5f6578e7c

- European Health Interview Survey wave 2 validation rules, Available from https://circabc.europa.eu/sd/a/c4686034-8d9e-4b96-93a2-5488b63db082/EHIS%20wave%202%20-%20Validation%20rules%20%282015-11-23%29.pdf.

- Emerging Risk Factors Collaboration, Di Angelantonio E, Sarwar N, Perry P, Kaptoge S, Ray KK, Thompson A, Wood AM, Lewington S, Sattar N, Packard CJ, Collins R, Thompson SG, Danesh J. Major lipids, apolipoproteins, and risk of vascular disease. JAMA. 2009 Nov 11;302(18):1993-2000.

- Davidsson G, Demarest S, Tolonen H, De Vitiis C, Primatesta P, Agafitei L. European Health Interview survey Task Force II. Report on guidelines and quality criteria for population health survey design and methods. Eurostat, 2009. Available at https://circabc.europa.eu/sd/d/0d4d38ff-fb62-4bf6-ba53-ed64063fc193/Task%20Force%20II%20report%20on%20survey%20design%20%28November%202009%29.pdf. Accessed on 22 January 2013

- de Koning L, Merchant AT, Pogue J, Anand SS. Waist circumference and waist-to-hip ratio as predictors of cardiovascular events: meta-regression analysis of prospective studies. European Health Journal. 2007:28(7):850-56.

- Florvall G, Basu S, Larsson A. Apolipoprotein A1 is a stronger prognostic marker than are HDL and LDL cholesterol for cardiovascular disease and mortality in elderly men. J Gerontol A Biol Sci Med Sci. 2006;61:1262-6.

- GBD 2015 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks,

1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016;388:1659-1724.

- Hanas R, John G, Consensus Committee. 2010 Consensus statement on the worldwide standardization of the hemoglobin A1C measurement. Diabetes Care 2010;33:1903-4.

- Lundqvist A, Mäki-Opas T. Health 2011 Survey – Methods. National Institute for Health and Welfare, Report 8/2016. Available at http://urn.fi/URN:ISBN:978-952-302-669-8.

- MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J, Abbott R, Godwin J, Dyeer A, Stamler J. Blood pressure, stroke, and coronary heart disease. Part 1. Prolonged differences in blood pressure: prospective observational studies corrected for regression dilution bias. Lancet. 1990:335:746-74

- Malnick SD, Knobler H. The medical complications of obesity. QJM. 2006;99(9):565-79.

- Mindell J, Biddulph JP, Hirani V, Stamatakis E, Craig R, Nunn S, Shelton N. Cohort profile: the health survey for England. Int J Epidemiol. 2012 Dec;41(6):1585-93. doi: 10.1093/ije/dyr199. Epub 2012 Jan 9.

- NCD Risk Factor Collaboration (NCD-RisC). Effects of diabetes definition on global surveillance of diabetes prevalence and diagnosis: a pooled analysis of 96 population-based studies with 331,288 participants. Lancet Diabetes Endocrinol. 2015;3:624-37.

- Primatesta P, Chaudhury M. Core Module and additional topics. In Tolonen H. et al. (eds.) Recommendations for the Health Examination Surveys in Europe. Kansanterveyslaitoksen julkaisuja 2008:B21/2008. Available at http://http://urn.fi/URN:NBN:fi-fe201204193358. Accessed on 22 January 2013

- Reinivuo H, Hirvonen T, Ovaskainen ML, Korhonen T, Valsta LM. Dietary survey methodology of FINDIET 2007 with a risk assessment perspective. Public Health Nutr. 2010 Jun;13(6A):915-9. doi: 10.1017/S1368980010001096.

- Scheidt-Nave C, Kamtsiuris P, Gößwald A, Hölling H, Lange M, Busch MA, Dahm S, Dölle R, Ellert U, Fuchs J, Hapke U, Heidemann C, Knopf H, Laussmann D, Mensink GB, Neuhauser H, Richter A, Sass AC, Rosario AS, Stolzenberg H, Thamm M, Kurth BM German health interview and examination survey for adults (DEGS) - design, objectives and implementation of the first data collection wave. BMC Public Health. 2012 Sep 1;12:730. doi: 10.1186/1471-2458-12-730.

- Seidell JC, Després J-P, Bouchard C. Waist and hip circumference have independent and opposite effects on cardiovascular disease risk factors: The Quebec Family Study. ACJN 2001;74:315-21.

- Sherwani SI, Khan HA, Ekhzaimy A, Masood A, Sakharkar MK. Significance of HbA1c Test in Diagnosis and Prognosis of Diabetic Patients. Biomark Insights. 2016;3;(11):95-104.

- Sierra-Johnson J, Fisher RM, Romero-Corral A, Somers VK, Lopez-Jimenez F, Ohrvik J, Walldius G, Hellenius ML, Hamsten A. Concentration of apolipoprotein B is comparable with the apolipoprotein B/apolipoprotein A-I ratio and better than routine clinical lipid measurements in predicting coronary heart disease mor-

tality: findings from a multi-ethnic US population. Eur Heart J. 2009;30:710-7.

- The International Expert Committee. International expert committee report on the role of the A1C assay in the diagnosis of diabetes. Diabetes Care 2009;32:1327-34.

- Tolonen H. et al. (eds.) Review of Health Examination Surveys in Europe. Publications of the National Public Health Institute 2008:B18/2008. Available at http://urn.fi/URN:ISBN:978-951-740-843-1. Accessed on 22 January 2013

- Tolonen H. Towards the high quality of population health surveys. Standardization and quality control. Publications of the National Public Health Institute 2005:A27/2005. Available at http://urn.fi/URN:NBN:fi-fe201204193557. Accessed on 22 January 2013

- Tolonen H, Koponen P, Mindell JS et al. Under-estimation of obesity, hypertension and high cholesterol by self-reported data: comparison of self-reported information and objective measures from health examination surveys. Eur J Public Health 2014; 24(6): 941-8

- Vartiainen E, Laatikainen T, Peltonen M, Juolevi A, Männistö S, Sundvall J, Jousilahti P, Salomaa V, Valsta L, Puska P. Thirty-five-year trends in cardiovascular risk factors in Finland. Int J Epid 2009: 1-15

- WHO. Waist Circumference and Waist-Hip Ratio: Report of a WHO Expert Consultation, Geneva, 8-11 December 2008. World Health Organization, 2011a. Available from http://apps.who.int/iris/bitstream/10665/44583/1/9789241501491_eng.pdf?ua=1

- WHO. Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus. Abbreviated Report of a WHO Consultation. World Health Organization, 2011b. Available from http://www.who.int/diabetes/publications/report-hba1c_2011.pdf

- WHO. Action Plan for implementation of the European Strategy for the Prevention and Control of Noncommunicable Diseases. World Health Organization, Regional Office for Europe, 2012. Available from http://www.euro.who.int/en/publications/abstracts/action-plan-for-implementation-of-the-european-strategy-for-the-prevention-and-control-of-noncommunicable-diseases-20122016

- WHO. Global action plan for the prevention and control of noncommunicable dieseases 2013-2020. World Health Organization, 2013.
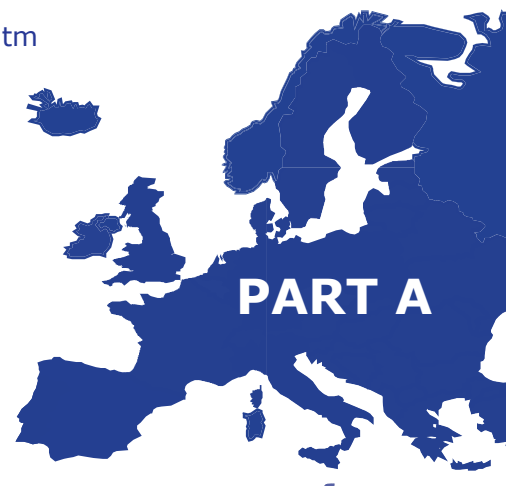
6. Timing of the fieldwork and order of measurements

PART A

Päivikki Koponen[1], Hanna Tolonen[1], Laura Paalanen[1], Johan Heldal[2]

[1] National Institute for Health and Welfare (THL), Helsinki, Finland
[2] Statistics Norway (SSB), Oslo, Norway

This chapter describes general issues related to timing of the field-work and order of measurements that need to be taken into account when planning the survey. For example seasonal and diurnal variation in symptoms, morbidity, body functions and health behavior need to be taken into account. Timing of the examinations will also affect participation rates (see Part A, Chapter 13.). General principles for the order of measurements need to be considered when estimating the time needed to carry out the fieldwork. These issues will all have an effect on personnel resources and other survey costs (see Part A, Chapter 16.). Details for timing specific measurements will be given in Part B of the EHES Manual.

6.1 Periodicity

The recommendation is to repeat the national HES with the EHES core measurements every five or six years, depending on national and European needs (e.g. if the EHIS and EHES data collection is combined, the periodicity needs to be in line with the regulation for EHIS). Some additional measurements may be repeated less frequently More frequent surveys do usually not reveal interpretable changes for the EHES core measurements, but they can be considered if there is a need to closely follow trends related to potential effects of specific health promotion activities.

An alternative is to build a system of continuous data collection. In such surveys data from different years can be aggregated to provide precise estimates for indicators that require larger samples. Continuous data collection also allows keeping permanent fieldwork staff, which may decrease staff recruitment and training costs. Examples of HESs with continuous data collection are the Health Survey for England and the National Health and Nutrition Examination Survey of USA. In a continuous survey the permanent core survey content can be kept brief while varying additional measurements can be introduced yearly or every second year. However, more comprehensive survey content will

allow more possibilities to study how different health topics are related to each other at individual level. The feasibility of a short or a longer, more comprehensive examination may vary between countries.

6.2 Length and time of year for the fieldwork

High seasonal variation has been identified in several health determinants as well as in biological measures. For example, seasonality impacts physical activity patterns (Merchant et al 2007, O'Connell 2014, Pivarnik et al 2003), food consumption (Fowke et al 2004, Locke et al 2009, Stelmach-Mardas et al 2016) as well as quality of life (Jia et al 2009). In countries and regions with cold winters, leisure-time physical activity is more common during summer and spring than during winter and autumn (Merchant et al 2007, Pivarnik et al 2003). People may be more likely to eat fruits and vegetables (Fowke et al 2004, Stelmach-Mardas et al 2016) and to report better quality of life (Jia et al 2009) during summer than during winter. Significant summer-winter differences have been identified in blood pressure, fasting plasma glucose levels, blood lipid levels, body mass index and waist circumference, with lowest risk factor levels in the summer (Chen et al 2006, Marti-Soler 2014, Visscher & Seidell 2004).

For the estimation of trends repeated surveys need to be carried out at the same time of the year. For best international comparability, the surveys in each country should cover evenly all seasons, which means that the fieldwork should last at least a year. Seasonal variation may differ between countries and regions, depending on the climate. If the survey covers only a part of the year, it is essential to evaluate the potential effect of weather and national/regional climate and other issues related to fieldwork timing (e.g. common flu epidemics) to measurement results as well as to participation rates.

Short survey duration usually needs a relatively large temporary staff, whereas long or yearly repeated surveys allow a more stable employment of the core staff. When the survey lasts more than a few months, particular attention needs to be paid to regular quality control, re-testing and re-training of the fieldwork staff, as well as to potential staff turnover during the data collection.

To identify health inequities comparisons by gender, age, socio-economic status and ethnicity, as well as by geographical regions will be possible only if the examinations in all such population subgroups are distributed evenly over the whole survey period. Regional comparability needs to be taken into account when scheduling visits to the different examination sites. A safe option is to order the examination sites of the country in random or to employ several teams to carry out the fieldwork simultaneously even though this may create some logistic challenges and increase the cost. Alternatively, systematic ordering where each region is visited evenly in all seasons can be used.

6.3 Weekdays and time of day

To allow easy access for participants and to minimize the effect of timing to measurement results, morning, day and evening appointments should be available, as well as several weekdays (see Part A, Chapter 13.). Also weekends should be used, if it is feasible to schedule these from the point of view of cost and availability of premises and staff, and if they are preferred by the participants. Measurements that require overnight fasting may be organized only in the mornings and may therefore be feasible only for a subsample (see Part A, Chapter 10.). Practical undocumented experiences from previous surveys in several countries have shown that the working age population prefers early mornings (before working hours) or late afternoons and early evenings (after work) for their examinations during the week. Fridays also seem to be less often preferred than other days of the week. In some surveys additional options for an appointment during weekends (Saturdays) have been used to raise willingness to participate.

Health behaviours, such as dietary habits vary between weekdays and weekends (An 2016, Yang 2014). Moreover, many of the HES measurements, such as blood pressure and some blood analyte concentrations have diurnal variation. Blood pressure tends to be lowest at night with a subsequent morning blood pressure surge (Kario 2010, Kario 2016). For lipids, it is difficult to dissociate the changes in their concentration from the effects of a meal. The most apparent postprandial changes have been observed for serum triglycerides (Sabaka 2013). The serum triglyceride level gradually increases after a meal, reaches a peak at 3–4 hours after the meal, and then slowly returns to its initial level at 6–8 hours after the meal (Miyosho 2014, Yunoki 2011). For various cholesterol fractions, postprandial changes occur but they are of smaller magnitude (Sabaka 2013, Wojczynsk 2011). Blood glucose concentration also rises after meals (American Diabetes Association 2001). However, blood glucose varies also irrespective of meals with a rise during sleep (Van Cauter 1991). Given the diurnal and postprandial variation, it is important to record the length of fasting and the time of the day when these measurements are performed (see Part B of the EHES Manual).

6.4 Order of measurements

The order of measurements has often constraints because of logistical requirements, such as composition of the fieldwork teams, study protocol (subject flow), costs and the examination visit's length. However, the following requirements need to be taken into account to ensure valid measurements and comparability between surveys.

6.4.1 Clinical measurements

The order of measurements should be determined as much as possible by (adapted from Tolonen et al 2002, Tolonen et al 2008):

1. Importance of the measurement; most important measurements should be done early in the session, in case the

participant is unable to follow the full examination protocol (time constraints, limitations in functional capacity etc.).

2. Sensitivity of questions and measurements; uncontroversial questions and measurements shouldn't be introduced first to allow building trust between the interviewer/measurer and the participant, but occur early or in the middle of the protocol to allow participants to become relaxed and comfortable with the procedures.

3. Stressfulness of procedure; blood pressure measurement should precede venepuncture and other potentially (mentally or physically) stressful tests/interviews.

4. Order in previous surveys; unless there are good reasons for change, it is recommended to maintain the former order of measurements to avoid bias.

5. Other effects on measurement results; blood pressure and blood samples should be taken before physical fitness tests or tests of physical function.

The following order is recommended for the EHES core measurements: blood pressure first, anthropometric measurements second, blood samples third and all additional measurements after these.

6.4.2 Questionnaires and interviews

The selection of self-administered questionnaires and interviews is described in Chapter 8. To avoid respondent burden, the questionnaire data collection may be split into parts administered before, during and after examinations. The decisions on when the questionnaires or interviews will be administered should be based on the following:

1. **Before the examination**

   Paper questionnaires and/or a link to a web based questionnaire can be mailed with the invitation to examinations, given or interviewed at a separate interviewer visit or phone interview, or administered at the examination site before the physical measurements.)

   When the questionnaire data is collected before the examination, the responses are not affected by the examination. Self-administered questionnaires should be checked at the examination site to avoid missing data or confusing responses. To promote participation in the examinations, it is recommended that the mailed (or web-based) questionnaires are easy to fill in and limited to most important key questions and not including questions that can be considered too sensitive. The core EHES questions (presented in Part A, Chapter 5.) are recommended to be administered before the examinations.

   If a separate interview phase is conducted before the health examinations, it's recommended that also the interview nonparticipants get the invitation to the examinations to avoid cumulating non-response. The time lag between the

interviews and the examination should be as short as possible to allow the use of interview data for the EHES indicators (e.g. medication as criteria for hypertension). If the interviewers have first contact with the selected persons, they can be trained to motivate participation to examinations and to book a time for the examination visit that best suits the participant.

2. **During the examination (between measurements)**

   When the self-administered questionnaires/interviews are completed during the examination visit, the responses can be affected by the measurements (learning that the participant has e.g. high blood pressure, knowing that smoking behaviour can be detected from blood/saliva cotinine etc.). It is recommended to ask questions on acute symptoms and current medication during the examination as these may affect the physical measurements.

3. **After the examination**

   Additional questionnaire data can be collected after the examinations, e.g. information on sensitive questions and questions that are less important from the point of view of the key aims of the survey, but which potentially raise new issues for research and health policy or health care development. Purposes for collecting sensitive information should be explained to the participants at the examination site to ensure that these questions do not worsen the participant's final impression of the survey.

References

- American Diabetes Association. Postprandial blood glucose. Diabetes Care 2001;24:775-778.
- An R.Weekend-weekday differences in diet among U.S. adults, 2003-2012. Ann Epidemiol. 2016;26(1):57-65.
- Chen S-H, Jen I-A, Chuang S-Y, Lin K-C, Chou P. Community-based study on summer-winter differences of component of metabolic syndrome in Kinmen, Taiwan. Preventive Medicine 2006;43:129-135.
- Fowke JH, Schlundt D, Gong Y, Jin F, Schu XO, Wen W, Liu DK, Gau YT, Zheng W. Impact of season of frood frequency questionnaire administration on dietary reporting. Ann Epidemiol 2004 14(10):778-785.
- Jia H, Lubetkin EI. Time trends and seasonal patterns of health-related quality of life among U.S. adults. Public Health Rep 2009;124(5):692-701.
- Kario K. Morning surge in blood pressure and cardiovascular risk: evidence and perspectives. Hypertension 2010;56:765–773.
- Kario K. Perfect 24-h management of hypertension: clinical relevance and perspectives. J Hum Hypertens. 2016 Sep 8. doi: 10.1038/jhh.2016.65. [Epub ahead of print]
- Locke E, Coronado GD, Thompson B, Kuniyuki A. Seasonal variation in fruit and vegetable consumption in a rural agricultural community. J Am Diet Assoc 2009; 109(1):45-51.

- Miyoshi T, Noda Y, Ohno Y, Sugiyama H, Oe H, Nakamura K, Kohno K, Ito H. Omega-3 fatty acids improve postprandial lipemia and associated endothelial dysfunction in healthy individuals – a randomized cross-over trial. Biomed Pharmacother 2014;68:1071–1077

- Marti-Soler H, Gubelmann C, Aeschbacher S, Alves L, Bobak M, Bongard V, Clays E, de Gaetano G, Di Castelnuovo A, Elosua R, Ferrieres J, Guessous I, Igland J, Jørgensen T, Nikitin Y, O'Doherty MG, Palmieri L, Ramos R, Simons J, Sulo G, Vanuzzo D, Vila J, Barros H, Borglykke A, Conen D, De Bacquer D, Donfrancesco C, Gaspoz JM, Giampaoli S, Giles GG, Iacoviello L, Kee F, Kubinova R, Malyutina S, Marrugat J, Prescott E, Ruidavets JB, Scragg R, Simons LA, Tamosiunas A, Tell GS, Vollenweider P, Marques-Vidal P. Seasonality of cardiovascular risk factors: an analysis including over 230 000 participants in 15 countries. Heart 2014;100(19):1517-23.

- Merchant AT, Dehghan M, Akhtar-Danesh N. Seasonal variation in leisure-time physical activity among Canadians. Can J Public Health 2007:98(3):203-208.

- O'Connell SE, Griffiths PL, Clemes SA. Seasonal variation in physical activity, sedentary behaviour and sleep in a sample of UK adults. Ann Hum Biol 2014 41(1):1-8.

- Pivarnik JM, Reeves MJ, Rafferty AP. Seasonal variation in adult leasue-time physical activity. Med Sci Sports Exerc 2003 35(6):1004-1008.

- Sabaka P, Kruzliak P, Gaspar L, Caprnda M, Bendzala M, Balaz D, Oravec S, Dukat A. Postprandial changes of lipoprotein profile: effect of abdominal obesity. Lipids Health Dis. 2013 Dec 8;12:179. doi: 10.1186/1476-511X-12-179.

- Stelmach-Mardas M, Kleiser C, Uzhova I, Peñalvo JL, La Torre G, Palys W, Lojko D, Nimptsch K, Suwalska A, Linseisen J, Saulle R, Colamesta V, Boeing H. Seasonality of food groups and total energy intake: a systematic review and meta-analysis. Eur J Clin Nutr 2016 70(6):700-708.

- Tolonen H, Kuulasmaa K, Laatikainen T, Wolf H, for the European Health Risk Monitoring Project. Recommendations for indicators, international collaboration, protocol and manual of operations for chronic disease risk factor surveys. 2002 [cited 13 2008]; Available from: http://www.thl.fi/publications/ehrm/product2/title.htm. Accessed on 22 January 2013

- Tolonen H, Koponen P, Aromaa A, et al. (Eds.) Recommendations for the Health Examination Surveys in Europe. B21/2008, Publications of the National Public Health Institute, Helsinki 2008. Also available from http://urn.fi/URN:ISBN:978-951-740-838-7. Accessed on 22 January 2013

- Van Cauter E, Blackman JD, Roland D, Spire JP, Refetoff S, Polonsky KS. Modulation of glucose regulation and insulin secretion by circadian rhythmicity and sleep. J Clin Invest. 1991;88:934-42.

- Visscher TLS, Seidell JC. Time trends (1993-1997) and seasonal variation in body mass index and waist circumference in the Netherlands. International Journal of obesity 2004;28:1309-1316.

- Wojczynsk MK, Glasser SP, Oberman A, Kabagambe EK, Hopkins PN, Tsai MI, Straka RJ, Ordovas JM, Arnett DK. High-fat meal effect on LDL, HDL, and VLDL particle size and number in the Genetics of Lipid-Lowering drugs and diet network (GOLDN): an interventional study. Lipids Health Dis. 2011, 10: 181- 10.1186/1476-511X-10-181.

- Yang PH, Black JL, Barr SI, Vatanparast H. Examining differences in nutrient intake and dietary quality on weekdays versus weekend days in Canada. Appl Physiol Nutr Metab. 2014;39:1413-7.

- Yunoki K, Nakamura K, Miyoshi T, Enko K, Kohno K, Morita H, Kusano KF, Ito H. Ezetimibe improves postprandial hyperlipemia and its induced endothelial dysfunction. Atherosclerosis 2011;217:486–491.

7. Selecting the examination site

PART A

Johanna Mäki-Opas[1], Hanna Tolonen[1], Päivikki Koponen[1], Kari Kuulasmaa[1]

[1] National Institute for Health and Welfare (THL), Helsinki, Finland

This chapter outlines the possible examination sites and their advantages and disadvantages. The selection of the survey site should be based on general requirements, national practices and cultural factors. Examination site may have an impact on the quality of the data and participation rate. The choice of the examination site will also depend on the availability of personnel (see Part A, Chapter 9.) and the national/regional health care system (e.g. possibilities for collaboration between the research institute and the health care organizations).

Potential examination sites are:

- Examination site within existing health care premises, such as a health centre or general practitioner's (GP) office
- Temporary examination site outside health care organizations, for example school premises, community centres or town halls;
- Mobile examination site, for instance a bus equipped for examination.
- Participant's home

7.1 Requirements for examination site

When physical examinations take place somewhere else than in the participant's home, the following issues should be considered:

- Participants should have easy access to the examination site. The maximum distance to the examination site varies between countries and even between areas within countries. In urban areas, people may not be willing to travel to another side of the city, but in rural areas longer distances can be considered acceptable.

- Availability of public or in other transportation to the examination site needs to be assured.
- Access of participants with limited functional ability (e.g. preferably no steep stairs in the building);
- Handling and storage of the blood samples;
- Requirements for the EHES core measurements;
  - Privacy;
  - Quietness;
  - Comfortable room temperature;
- Requirements for the additional measurements, e.g.
  - Enough space for functional ability tests;
  - Sound proof environments for audiograms.

The only way to be sure that the examination site is suitable for carrying out physical measurements is to visit the place before selecting it. This may require extra time and personnel resources during survey preparation.

7.2 Requirements for home visit

Home is a private place. When the examinations takes place at participants home some special issues should be taken into account.

These are for example:

- Acceptability of home visits among the population, e.g. are people used to home visits by the primary health care personnel;
- Special attention to the safety of the fieldwork staff should be paid (e.g. safety while travelling, walking in the neighbourhood and if alone in the house with a temperamental participant);
- It may be difficult to guarantee privacy during the examinations/interview if family members are present;
- Standardization and calibration of the equipment and following the measurement protocol may be challenging;
- Restrictions for handling and storage of the blood samples may compromise the quality of the samples;
- Challenges for data transfer and data confidentiality.
- Measurements that need equipment that is heavy or otherwise difficult to transport or have other special requirements for the environment, cannot be conducted at home;

When other examination sites are used, it should be considered if it is feasible and useful to offer home visits to those who are not willing or able to come to the examination site (e.g. due to limited functional ability).

7.3 Advantages and disadvantages of different examination sites

All examination sites have their advantages and disadvantages (Table 7.1).

Table 7.1 Advantages and disadvantages of the examination sites

	Participant's home	Temporary examination site	Examination site within existing health care premises	Mobile examination site
Access by partici-pants	Easy access	Requires effort	Requires effort	May be easy if mobile ex-amination site can be taken close to the par-ticipants
Cost for partici-pants	None	Travel costs	Travel costs	Some travel costs
Environment				
Atmosphere	Relaxed	Some tension	Some or a lot of ten-sion	Some ten-sion
Privacy	Limited pri-vacy if other family mem-bers at home	Can be con-trolled	Can be controlled	Can be controlled
Temperature	Cannot be controlled by the survey team	Can usually be controlled	Can usually be con-trolled	Can be controlled
Quietness	Cannot be controlled by the survey team	Can usually be controlled	Can usually be con-trolled	Can be controlled
Safety of the fieldwork staff	Cannot be controlled	Can be con-trolled	Can be controlled	Can be controlled
Travel cost of fieldwork staff	Expensive	Some	Some	Some
Traveling for field-work staff	Lot of trave-ling	Some	Some	Some

	Participant's home	Temporary examination site	Examination site within existing health care premises	Mobile examination site
Restriction to measurements	Only measurements for which devices can be transported easily and which do not have specific environmental requirements	Generally none	Generally none	Generally none, sometimes a lack of facilities for specific measurements may come up (e.g. limited space)
Calibration/standardization of the measurements	Difficult	Can be done	Can be done (but if equipment of health care centre are used, standardization and calibration may be difficult)	Can be done
Acceptability	Some people are not willing to let the survey team into their home	Generally accepted	In some countries there may be differences among population groups which organizations are valued. This may affect willingness to participate.	Generally accepted
Time and cost for setting up an examination site	Minimal	Time consuming	Takes some time (depends on used equipment, if equipment from the health care facilities are used, carefull calibration before fieldwork is needed, otherwise like temporary examination site)	Time consuming and costly
Cost of the maintenance of the examination site	None	Some costs	Some costs (depends on agreements with the local health care administration)	Costly

8. Questionnaire design and administration

PART A

Päivikki Koponen[1], Johanna Mäki-Opas[1], Hanna Tolonen[1]

[1] National Institute for Health and Welfare (THL), Helsinki, Finland

This chapter covers issues that need to be taken into account when preparing the national HES questionnaire and planning the questionnaire administration. The questionnaire design has an impact on participation rate and validity of the obtained data. The questionnaire administration mode has an effect on survey budget but it may also have an effect on participation rate, item non-response and validity of the answers.

8.1 Questionnaire design

Every national HES should include a questionnaire to collect information which is needed e.g. to interpret the measurement results. The questionnaire design affects the participation rate as it gives the participant an impression of how easy, convenient and time-consuming it is to take part in the survey. It also has an effect on reliability and accuracy of the obtained information. Therefore, enough time and resources should be allocated for planning the questions and preparing the questionnaire (Tolonen 2005). Usually the questionnaire is at least slightly modified and improved after the experience from the pilot survey.

Language, wording of the questions, selection of the response alternatives, formulation of sensitive questions, prevention of recall bias, order of questions, jump rules and the length of the questionnaire are the main elements of questionnaire design.

8.1.1 Language and wording

Proper wording of the questions is essential; the questions should be simple and straightforward. This helps to ensure that respondents understand the questions correctly. Effort must be devoted to avoiding ambiguity in the wording of the questions. Professional or highly technical terms, slang, abbreviations or words which may be considered

as insulting should be avoided. In each question only one issue should be addressed. All questions should be available in the native language of the respondent. (Rea 2005.) In many European countries, several language versions should be considered if there are several official languages in the country or if there are major ethnic minorities with poor language skills in the national language(s).

The translations should be prepared with a careful validation process. The translator should preferably be a health professional, familiar with terminology in the questionnaire and with interview skills. The focus of questionnaire translation should be on cross-cultural and conceptual, rather than on linguistic/literal equivalence (WHO). Previous experience shows that a scientific back-translation process isn't always feasible and not even necessary (Font & Mendez 2013). The EHES core questionnaire (see Part B, Chapter 5 of of the EHES Manual) consists many questions from the European Health Interview Survey (EHIS). The EHIS questions have usually been translated to the national language(s) by the national Statistics Office and at least for some sets of questions, cognitive validation has been done. These EHIS translations should be used whenever possible.

8.1.2 Recall bias

When formulating the questions it is good to remember that people tend to forget events. It is usually easier to remember things that happened recently than for example a year ago. When the recall period is longer the accuracy is often worse. Recall can become a source of bias (de Bruin 1996). Recall of events can be assisted by adding aids to the questionnaire and by ordering of the questions. For example holidays and national festivals can be used to refer to a certain time period, or the respondents can use a calendar. (Tanur 2004)

8.1.3 Order of the questions

The order of the questions in the questionnaire is also important. A poorly organized questionnaire may confuse the respondent, bias the responses, and have an effect on the response rate, as well as on the willingness to answer sensitive questions. (Rea et al. 2005, Tanur 2004, Biemer et al. 1991.) The questionnaire should start with the easy questions. When more difficult questions are placed at the end of the questionnaire and the respondent stops answering them, at least some data for earlier questions will be obtained. During the interview asking the easy questions first may help to build trust between the interviewer and the respondent so the respondent may be more willing to answer more difficult questions in the end.

The questions should be grouped by the topic. This makes answering easier and helps to reduce recall bias. Filtering questions should also be used. This reduces the respondents burden. Use of jump-rules in the questionnaire avoids respondents answering irrelevant questions. Also the order of the response alternatives can greatly influence the results (Biemer et al. 1991).

Each national HES should include at least the EHES core questions, preferably in the same order as in the EHES model questionnaire (see Part B, Chapter 5). If the national questionnaire includes several additional items, it is recommended to keep the EHES core questions early in the questionnaire to make sure that the participants give valid responses to all of them. However, the structure of the whole questionnaire needs to be taken into account.

8.1.4 Length of the questionnaire

The length of the questionnaire affects the response rate as well as reliability of the data. A short questionnaire increases the response rate but may lack questions for important indicators. With the longer questionnaire the respondents often get careless towards the end and the reliability of the answers suffers (Biemer et al. 1991). The ideal length for filling in a self-administered questionnaire is 15-30 minutes and for the face-to-face interview 30-60 minutes. In practice, questionnaires which are designed for these lengths, may require about 15 minutes longer for most respondents. (Rea et al. 1997)

8.1.5 Layout of the questionnaire

Issues to be considered when using paper questionnaires include e.g.:

- font size and font style feasible for persons with problems in visual capacity (especially if elderly persons are included);
- number of questions on each page;
- number of pages needed;
- if some questions can be skipped, clear advice to jump to next questions;
- using colours and pictures can help the respondent to focus, to choose the answers and understand the questions (e.g. food or alcohol portions).

Issues to be considered when using electronic/web questionnaires and when developing or choosing software for computer aided personal interviews (CAPI), computer aided telephone interviews (CATI) or computer aided self-interviews (CASI) include e.g.:

- possibility to choose language (if several languages are needed);
- font size and font style;
- number of questions visible at each screen;
- skipping of irrelevant items/sections which are conditional on the answers to previous questions, jump rules to be followed (controlled by the program, not the respondent);
- using colours and pictures;
- possibility to stop filling-in the questionnaire and continue later;
- automatic checking of data and consistency checks to minimize invalid responses and item non-response (e.g. pre-

set maximum and minimum values, and confirming missing data);

- downloading the file in respondent's own computer not too time consuming;
- functioning in both computers and mobile devices;
- data security.

8.2 Questionnaire administration

Survey questionnaires can be filled in either by the respondent (i.e. self-administration) or by an interviewer. Both self-administration and interview have several alternatives how they can be organized and all of them have advantages and disadvantages, see Table 8.1. (Franklin & Walker 2003, Czaja & Blair 2005, Tolonen 2005). The questionnaire administration mode may have an effect on participation rate and the accuracy and reliability of the responses. To improve general data quality and to avoid item non response it is recommended that the core EHES questions are collected through face-to-face interview or that the self-administered questionnaires are checked with the participant during the visit to health examinations, at least for the key questions. Other administration modes can be considered for additional questions and when the person does not respond to the first contact attempt or refuses to take part in the examination (for the non-respondent questionnaire).

Use of mixed-mode data collection and several phases of questionnaire administration may avoid participant's burden and selection bias. The combination of multiple modes may offer a means to improve overall survey response rates and possibly broaden population coverage (Sinclair et al 2012). There are also disadvantages in using several modes of data collection in one survey or in using different modes for different periods of data collection or different modes in different areas or countries, This may cause bias for comparisons between different population groups or countries as the mode of data collection affects the respondent's responses, especially for items which might be affected by social desirability or which are considered to address sensitive or highly personal issues (e.g. income, alcohol and substance use) (Okamoto et al 2002, Bowling 2005, Link & Mokdad 2005a). Also the characteristics of persons responding to different modes may differ, e.g. students, younger persons and those with higher education responding to mailed or web-questionnaires more often than others (Link & Mokdad 2005b).

Table 8.1 Comparison of different questionnaire administration methods (adapted from Franklin & Walker 2003, Bowling 2005, Czaja & Balir 2005, Tolonen 2005)

Aspect	Self-administered paper questionnaire	Self-administered electronic questionnaire	Interview Telephone	Interview Face-to-face
Cost	Low	Low*	Medium	High
Length of questionnaire (number of questions)	Medium	Short	Medium	Long
Complexity of questionnaire (e..g jump rules)	Must be simple	May be complex	May be complex	May be complex
Control of question order	Low	Fair	Very good	Very good
Use of visual aids	Good	Very good	Not possible	Very good
Feasibility of sensitive topics	Good	Low	Fair	Fair
Control of response situation	Low	Low	Fair	Good
Possibility to choose the language version	Low	Very good	Good	Very good
Socially desirable answers	No	No	Yes	Yes
Item non-response	High	High /Low**	Medium	Low
Response rate	Low	Low	Medium	High
Needed literacy level	High	High	Low	Low
Verifying the respondents identity	Low	Low	Medium	High

* Setting up the electronic/web questionnaire may be costly but after that costs of data collection are low
** Depending on the programme used

8.2.1 Self-administration

Self-administration of the questionnaire is cost effective but assumes that participants have a good literacy level. Persons with visual problems and other problems in functional ability may also need extra help to avoid selection bias. A self-administered questionnaire should be relatively short and all questions need to be completely self-explanatory; format and question wording must be simple. Self-administration eliminates the interviewer effect but may result in missing data as a result of uncertainty about the question. The self-administered questionnaire can be either a paper form or an electronic version. Paper forms require separate data entry. The electronic questionnaire can be at the internet or on stand-alone software on computer at the health examination site. The electronic questionnaire can be more complex (with skip patterns) than the paper format. The computer programme should have built-in checks for responses (e.g. upper or lower limits for response categories).

A self-administered questionnaire can be mailed with the invitation to be filled in at home before the examination and checked by field work staff at the examination site. The possibilities to motivate participation

to examinations are poor if questionnaires are mailed before examinations. It is also known that response rates tend to be low when self-administration is used. Alternatively, the questionnaire can be given to the participant when he/she arrives at the examination site and he/she fills in the questionnaire at the examination site. In this case, the participant can ask help from the field work staff if he/she has any problems with the questionnaire. Also in this case, the completed questionnaire should be checked by the field work staff for completeness before the participant leaves the site.

Self-administration provides more privacy for the respondent than interview, and it is particularly suitable for sensitive questions (e.g alcohol and substance use, sexual behavior, income). The questionnaire can contain printed reference materials and pictures (visual aids). For example, pictures can be useful for showing portions in questions on alcohol intake and food consumption/diet.

Web-based questionnaires can be easy for certain groups of the population. In most European countries they are more feasible as an alternative to the traditional paper forms, than as an exclusive mode of data collection. Also mobile phone survey applications can be considered.

8.2.2 Interviews

Interviews are time consuming and carry additional personnel costs, but they eliminate the issues of low literacy level and functional impairment and provide an opportunity for clarifying the questions if needed. These clarifications have to be described in the manual and training for the interviewers and/or in the questionnaires to avoid biased responses and to ensure standardization of questionnaire administration. Interviewer effects related to gender, ethnicity and other interviewer characteristics need to be considered as they may be more important in some countries and cultures than in others (Davis et al 2009).

Interviews can be conducted either by telephone or face-to-face. In both modes the questionnaires can be quite long and complex, if skip patterns and jump rules are used and followed by the interviewer and/or controlled by the computer programme. This reduces the burden of the respondent and controls the length of the interview. Automatic built-in checks for responses and data entry by the interviewer may reduce data errors in computer assisted interviews.

Face-to-face interviews are usually the most expensive mode for questionnaire administration. However, they have many advantages: the interviewer has a possibility to check the personal records (e.g. medication), personal contact may increase the response rate and the use of printed reference materials (visual aids) is possible. Telephone interviews are less expensive but provide no control over the environment in which the interview is conducted. Question wording needs to be simple. Telephone interviews also requires good hearing capacity from the respondent.

For all interviews, there is a risk that interviewers introduce bias by not asking the questions verbatim, modifying the questions or by incorrect

prompting. This risk can be reduced, but not fully eliminated, by proper training and quality control. Sensitive questions may be problematic in interviews, because the respondent may reply according to what is socially most acceptable.

8.2.3 Mixed mode

When there are several additional topics and many questionnaire items, a mixed mode should be considered: e.g. a short self administered questionnaire mailed before examinations, interview during examinations, and another questionnaire given to be filled in later at home. Several modes of data collection can also be used for the same questionnaire to obtain better response rates, e.g. self-administered questionnaires are mailed as a paper version to all subjects, but a possibility to fill this in as a web-based questionnaire is given in the cover letter. In addition, interviews during the examination may be offered to those who were unable to fill in the questionnaires by themselves. The mode of questionnaire administration should be recorded to allow comparison of responses by different administration modes.

8.3 Use of proxies

In EHES data collection, proxy use during the interviews is only allowed when the selected person him/herself is unable to respond due to major limitations in communication and/or cognitive ability. The reason for proxy use (why the selected individual was unable to respond on his/her own behalf) and type of proxy (spouse, child or other relative/ significant other, or nurse for e.g. institutionalized persons) should always be recorded. When the use of proxy is considered, special attention should be paid for the decision whether of not the person him/ herself is capable to provide informed consent (see Part B, Chapter 3. of the EHES Manual). Proxy use can be avoided by proper resources during data collection and scheduling adequate time to contact all selected persons.

There is a lot of evidence that the use of proxies introduces systematic biases, affecting national disability estimates and the incidence of several chronic conditions as well as their trends in repeated surveys (Shields 2004, Todorov & Kirchner 2000). Proxy responses and self-reports differ significantly depending on the type of questions, age and gender of both the proxy and the selected person, and relationship between the selected person and the proxy (Neumann et al 2000, Toldrov & Kirchner 2000, Shields 2004, Snow et al 2005). For younger persons there is evidence of proxy respondents under-reporting chronic conditions, disability and medication use, while for older persons the bias may be opposite, proxies reporting more impairment than self-respondents.

References

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

- Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health 2005 Sep;27(3):281-91.

- Czaja R, Blair J. Designing surveys. A Guide to Decisions and Procedures. 2nd edition, Sage Publications, Inc., California, USA, 2005.

- Davis, RE, Couper MP, Janz NK, Caldwell CH, Resnicow K. Interviewer effects in public health surveys. Health Educ Res. 2010 25(1):14-26.

- de Bruin A, Picavet HSJ, Nossikov A. Health interview surveys. Towards international harmonization of methods and instruments. WHO Regional Publications. European Series. 1996: No 58.

- Font J, Mendez M. Surveying Ethnic Minorities and Immigrant Populations. Amsterdam University Press, Amsterdam 2013.

- Franklin S, Walker C (eds). Survey methods and practices. Statistics Canada. Social Survey Methods Division, 2003.

- Link MW, Mokdad AH. Effects of survey mode on self-reports of adult alcohol consumption: a comparison of mail, web and telephone approaches. J Stud Alcohol. 2005a Mar;66(2):239-45.

- Link MW, Mokdad AH. Alternative modes for health surveillance surveys: an experiment with web, mail, and telephone. Epidemiology. 2005 b Sep;16(5):701-4.

- Neumann PJ, Araki SS, Gutterman EM. The use of proxy respondents in studies of older adults: lessions, challenges, and opportunities. J Am Ger Soc 2000; 48(12):1646-54.

- 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 May;55(5):505-11.

- Rea LM, Parker RA. Designing and conducting survey research. A comprehensive guide. San Francisco: Jossey-Bass; 2005.

- Shileds M. Proxy reporting of health information. Health Reports 2004; 15 (3): 21-33.

- Snow LA, Cook KF, Lin P-S, Morgan RO, Magaziner J. Proxies and Other External Reters: Methodological Considerations. Health Services Research 2005;40(5):1676-1693.

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

- Todorov A, Kirchner C. Bias in Proxies' Reports of Disability: Data from the National Health Interview Survey on Disability. American Journal of Public Health 2000; 90(8): 1248-1253.

- Tolonen H, Kuulasmaa K, Laatikainen T, Wolf H, for the European Health Risk Monitoring Project. Recommendations for indicators, international collaboration, protocol and manual of operations for chronic disease risk factor surveys. 2002. Available at http://www.thl.fi/publications/ehrm/product2/title.htm. Accessed on 22 January 2013

- Tolonen H. Towards the High Quality of Population Health Surveys. Standardization and Quality Control. National Public Health Institute A27/2005, Helsinki. Available at http://urn.fi/URN:ISBN:951-740-546-4

- WHO. Process of translation and adaptation of instruments. Available at http://www.who.int/substance_abuse/research_tools/translation/en/   . Accessed on 9 November 2016.

9. Selection of the fieldwork staff

PART A

Päivikki Koponen[1], Laura Paalanen[1], Hanna Tolonen[1]

[1] National Institute for Health and Welfare (THL), Helsinki, Finland

Competent and motivated fieldwork staff is a key to successful data collection. Characteristics of the staff members can influence nonresponse as well as validity and reliability of survey data, especially in public health surveys, which cover sensitive issues and topics prone to socially desirable responding (Davis et al 2010). In HES the selection of fieldwork staff has to be based on general requirements and competences needed to carry out the clinical measurements. Differences in the national health care systems as well as national guidelines need to be considered.

9.1 General principles and criteria for recruitment

Interviewer and measurer effects have to be considered when selecting and recruiting fieldwork personnel. Existing literature on interviewer race and ethnicity effects fails to conclude whether respondents feel more comfortable with, trust, prefer or provide more accurate data to interviewers of their own race, sex and ethnicity (Davis et al 2010). However, the possibility of such effects should be taken into account and evaluated. The general principles for the selection and recruitment of fieldwork staff are (adapted from Tolonen et al 2008):

1. Legislation concerning medical practice and nursing in each country as well as the EU directives (Directive 2005/36/EC amended by Directive 2013/55/EU) for the recognition of professional qualifications have to be taken into account.

2. The personnel should be motivated to strictly follow the survey protocols to ensure reliability and accuracy of the survey results.

3. General appearance (non-provocative, calm and neutral appearance and good manners), friendliness, respect, empathy, encouragement and interest shown towards par-

ticipants may affect participation and the results of the measurements. Age, gender, and ethnicity of the fieldwork personnel need to be taken into account in respect to the national and the participants' culture. It is recommended that the fieldwork teams consist of personnel with a variety of backgrounds. For example, in some cultures male nurses may not be accepted to carry out measurements requiring light clothing for women.

4. Willingness and possibility to travel around the country with the survey team may be needed depending on survey logistics. For example, this may be a problem for persons with small children.

5. Professional competence of the staff members and service given to participants may also be an important factor affecting survey response. Feedback given to the participants during and after the measurements needs to be considered also in the selection of survey staff. For example physiotherapists may be better qualified than nurses to carry out some physical functioning tests, while registered nurses may be better qualified than nurse assistants or a medical doctor to carry out blood pressure measurements.

6. Fluency in national language(s), and if needed, languages of the major migrant groups.

Fieldwork staff may be recruited specifically for the survey. An alternative is to use personnel from the local health care organizations (e.g. primary care units or health centers or hospitals) in the selected survey sites. It is usually easier to ensure standardization of measurements if fieldwork staff is recruited specifically for the survey. When permanent personnel of the local health services are trained to carry out the survey fieldwork they may be tempted to follow their regular practices instead of the survey protocols. This may happen especially if they also have their regular tasks during the survey, and are only part time carrying out the survey fieldwork. In any case the use of the local personnel in each survey site increases substantially the time and efforts needed for training. The use of regular health service personnel may also affect survey results as the familiarity may both enhance and restrict open communication.

A combination of the two groups of personnel may be considered. Specially recruited personnel travelling from survey site to another is trained to carry out the measurements that are most challenging to standardize, such as blood pressure measurements. These specific survey staff members may also supervise the local personnel responsible for other tasks. The local personnel may also be more efficient in recruiting participants.

9.2 Professional groups

Different professional groups have both benefits and disadvantages (Table 9.1). It is recommended that registered nurses carry out the EHES core measurements. The person performing the blood collection should be a certified phlebotomist. In most countries, this certification

is offered through national accrediting agencies for clinical laboratory sciences. Employing a certified phlebotomist for the invasive blood collection procedure provides not only comfort and safety for the participant but also some medical-legal protection for the survey organizers, in case something should go wrong. Other professional groups such as medical-technical assistants, nutritionists, dental assistants or physiotherapists can be considered

Table 9.1 Requirements, benefits and disadvantages of different professionals in survey fieldwork

Professional group	Specific requirements	Benefits	Disadvantages
Physicians (or dentists if oral health is measured)	Needed if clinical or diagnostic examinations are carried out and if physician's presence is required for clinical measurements according to national regulations.	May increase participation because of higher professional respect/regard among the population. Better readiness for acute situations during the fieldwork, and in interpreting test results and informing participants about their test results (better service to participants may affect willingness to participate).	High salary level (effect on survey costs). Higher tendency to adapt survey protocols, not follow standards (Graves & Sheps 2004) and make independent decisions (also in conflict with survey protocols). Higher "white coat"/observer effect on some measurements, such as blood pressure (Graves & Sheps 2004, Labinson et al 2008**).**
Nurse	Registered nurse generalists with training according to the Directive 2005/36/EC amended by Directive 2013/55/EU are recommended for most measurements	Better adherence to follow standards (Graves & Sheps 2004) in survey protocol than physicians. Lower salary level and lower survey costs compared to physicians.	Differences in professional independence and respect among the population in European countries.
Certified phlebotomists	Recommended for blood sample collection. In some countries this is a legal requirement.	In-depth qualifications for blood sample collection.	Differences in basic professional training in European countries.
Interviewers	Specific interviewer training recommended if personal (face-to-face) or telephone interviews are used.	Standardized interviewing techniques.	Interviewers with medical/nursing background are more qualified for asking questions on medical conditions and medication. Lay interviewers may get more valid answers to questions on health behavior, as people may tend to give more socially acceptable answers to professionals.

9.3 Fieldwork teams

When estimating the number of survey personnel needed for the fieldwork, potential sick leaves and other absences need to be anticipated. In most cases it is recommended to train a few extra persons for substitutes to ensure that time schedules can be kept and the participants are served as well as possible. Especially when the fieldwork period lasts for several months and the examinations are carried out by a team consisting of specific personnel for each measurement, the possibility to rotate duties between staff members should be considered. Such rotation of duties helps to minimize measurer effects and to motivate the staff members to follow the standards. This requires staff members with broader competence, who can also substitute other team members in case of absences (e.g. sick-leaves).

The number of fieldwork teams will depend on the length of the fieldwork period, and the distances between selected survey sites/locations. The effect of the number of teams and staff members on survey budgets can be estimated with the EHES budgeting tool (see Part A, Chapter 16 of the EHES Manual).

Support and supervision from the survey organizers or from the survey core group ("the central survey office") need to be ensured. This is particularly important if several teams work simultaneously in different survey locations. Well-defined leadership within the team is also essential. Each fieldwork team should have a specified supervisor/leader that follows the work progress and adherence to standards among all team members. Physicians may be needed to interpret measurement results or to give medical advice when abnormal measurement results are found and may need urgent consultation. When physicians are not part of the field teams their availability for consultation has to be organized in another way.

In case of home visits, the fieldwork teams seldom include more than two persons (interviewer and a nurse). They need to be well trained for making home visits. Typically public health nurses or health visitors are used as people most easily accept their visits. For surveys carried out in clinic settings the professionals selected for the fieldwork teams may vary. Two examples are presented here.

9.3.1 Example team for a survey in clinic environments and only the EHES core measurements

Nurse 1, tasks: reception of the participants, obtaining informed consent, short health interview or checking the self-administered questionnaire.

Nurse 2, tasks: Blood pressure measurement, height, weight and waist circumference measurement.

Phlebotomist, tasks: drawing and processing blood samples

Nurse 1 can be selected with less professional competence and with lower salary level (e.g. medical receptionist, medical-technical assistant). However, if nurse 1 and nurse 2 are both registered nurses, rotation of tasks e.g. with monthly intervals and substitution of the other nurse in case of sudden absences is possible. A survey physician may be needed as a back-up person (on call), easily available for consultation. This consulting physician can cover several fieldwork teams working in different locations.

9.3.2 Example team for a survey in clinic environments and also several additional measurements

Nurse 1, tasks: reception of the participants, obtaining informed consent, short health interview or checking the self-administered questionnaire.

Nurse 2, tasks: Blood pressure measurement, height, weight and waist circumference measurement, lung function test (spirometry).

Phlebotomist/bioanalyst, tasks: drawing and processing blood samples

Nurse 3, tasks: diagnostic mental health interview (e.g. CIDI)

Physiotherapist, tasks: hand grip strength measurement, test of standing balance and timed chair stand test

Physician, tasks: clinical medical examination with e.g. auscultation of the heart and lungs, interpreting previous measurement results (e.g. spirometry), and diagnostic assessments

In this team it is possible to rotate tasks between several team members if the bioanalyst is trained also to cover the tasks of nurse 2 and nurse 2 is also certified/qualified to draw blood samples. Nurse 1 and nurse 3 can easily be trained for both tasks. The last professional whom the participants meet at the end of the examination is the physician who checks all measurement results and may advice the participants to seek further medical help when needed.

References

- Davis RE, Couper MP, Janz NK, Calwell CH, Resnicow K. INterviewer effects in public health surveys. Health education research 2010;25:14-26.
- Directive 2005/36/EC of the European Parliament and of the Council of 7 September 2005 on the recognition of professional qualifications. *Official Journal of the European Union* 2005:22-142.
- Directive 2013/55/EU of the European Parliament and of the Council of 20 November 2013 amending Directive 2005/36/EC on the recognition of professional qualifications and Regulation (EU) No 1024/2012 on administrative cooperation through the Internal Market Information System ('the IMI Regulation'). Official Journal of the European Union 2013;56:132-170.

- Graves JW, Sheps SG. Does Evidence-based medicine suggest that physicians should not be measuring blood pressure in the hypertensive patient? Am J Hypertension 2004;17:354-360.

- Labinson PT, Giacco S, Gift H, Mansoor GA, White WB. The importance of the clinical observer in the development of a white-coat effect in African-American patients with hypertension. Blood Press Monit 2008 Jun;13(3):139-142.

- Tolonen H, Koponen P, Aromaa A, et al. (Eds.) Recommendations for the Health Examination Surveys in Europe. B21/2008, Publications of the National Public Health Institute, Helsinki 2008. Available at http://urn.fi/URN:ISBN:978-951-740-838-7. Accessed on 22 January 2013

10. Blood samples and laboratory analyses

PART A

Georg Alfthan[1], Jouko Sundvall[1]

[1] National Institute for Health and Welfare (THL), Helsinki, Finland

This chapter gives guidelines and describes requirements for selection of laboratories, collection of biological samples, processing and storage, and the laboratory analysis of the samples. Also procedures on quality requirements of analytical laboratories and guidelines for the standardization of methods are are described.

10.1 Selection of analytical laboratories

It is recommended that all analyses pertaining to the core measurements of a country should be performed at the same laboratory, hereafter called the National HES Laboratory (NHESL). The most important criteria for selection of the laboratory should be based on its performance in external quality assessment (EQA) programmes. Whenever possible, the laboratory should be accredited by a national organization.

Accreditation

A prerequisite for a laboratory to become accredited is to have a documented quality management system. The usual contents of the quality manual follow the outlines of either the ISO/IEC 17025 for Testing and Calibration Laboratories or the ISO 15189:2007 for Medical Laboratory Standards.

Laboratories use the above standards to implement a quality system aimed at improving their ability to consistently produce valid results. This is the basis for accreditation from a national Accreditation Body. Since the standard is about competence, accreditation is simply formal recognition of fulfillment of that competence.

In these instructions we assume that the NHESL will be responsible also for the long-term storage of all samples. If this is not the case in the country, it should be taken into account in the national HES Manual. A Central EHES Reference Laboratory (EHES RL) for support and EQA of the National HES Laboratories has been proposed and its availability is dependent of the future resources of the EHES RC.

10.2 Selection of measurements on the biological samples

Total cholesterol, HDL cholesterol and fasting glucose are core measurements, which should be included in all surveys. It is recommended that all countries collect more blood samples than are needed for the core measurements. This will make it possible to do various additional measurements on the samples in the future. Also collection of other biological samples is recommended if feasible. There is more discussion on potential additional measurements in Part A, Chapter 5.

For the time being, instructions for the analytical laboratory covers mainly the core measurements.

10.3 Blood collection

The collection of blood samples for the analysis of the core measurements is described here.

10.3.1 Core blood measurements

Total cholesterol and high density lipoprotein (HDL) cholesterol:
These lipids should be measured from serum. Fasting is not necessary.

Glucose:
Plasma glucose should be measured from fluoride-citrate plasma. 8-14 hours fasting is necessary. Because of potential difficulties in requiring fasting from all participants, the measurements may cover only a subsample of the survey.

- In case fluoride-oxalate or another agent is used as an anticoagulant/inhibitor, a 5% lower glucose concentration per each 30 min may be expected before separation of red cells.

Additional measurements are considered in Part A, Chapter 5 of the EHES Manual.

10.4 Critical issues of the blood collection

10.4.1 Fasting before the sample collection

The serum samples for total cholesterol and HDL cholesterol can be taken at any time of the day, with the subject non-fasting. If measuring fasting glucose, lipoprotein fractions and triglycerides, the samples should be collected after a fasting period of at least 8 hours and at most 14 hours (excessively long fasting causes major changes in energy metabolism, with implications for blood triglycerides). In practice, this means that fasting must be overnight and that the samples can only be taken in the morning and can only be expected from persons who are invited to undergo the examination in the morning (see also Part A, Chapter 6.). In all cases the length of time from the last meal in full hours should be documented.

10.4.2 Position of the subject

All blood samples should be drawn with the subject in a sitting position preceded by a 10-15 min rest. Preferably, blood should not be collected from the arm that is used for blood pressure measurement, (i.e., blood should usually be drawn from the left arm).

10.4.3 Use of a tourniquet

Prolonged venous occlusion can cause changes in concentrations of blood constituents. Therefore, the use of a tourniquet should be minimized. If a tourniquet is used to search for a vein, it should be released before withdrawal of blood begins. In any case, the use of a tourniquet should be limited to less than one minute.

10.4.4 Effects of seasonal variation

Diurnal effects on analyte concentrations are varied. For the lipids it is difficult to dissociate changes in their concentration from the effects of a meal. Studies suggest that cholesterol concentrations are higher in autumn and winter than in spring and summer.

10.4.5 Effect of physical training

Excessive physical training may cause dehydration resulting in raised serum electrolytes and several enzymes of muscle origin. Except for dehydration, other effects are difficult to estimate. Therefore, abstaining from heavy physical training for 8 hours preceding phlebotomy is recommended.

10.5 Equipment for drawing of blood samples

10.5.1 Choice of type and order of blood tubes

The number and type of blood collection tubes depend on the core and other anticipated measurements on the samples. The blood collection kit, including all tubes and equipment needed for the procedure, needs to be planned and prepared in such a way that all parts are compatible. An example of a kit, covering the measurements specified in section "Selection of measurements" above is provided in Table 10.1.

Table 10.1. A recommended kit including all supplies for blood collection, processing and storage

a -	plain serum gel tube (9/8 ml) used for core measurements, e.g. lipids, lipoproteins (serum)
b -	fluoride-citrate (5/3 ml) used for glucose, clotting factors, adhesion molecules (plasma)
c -	EDTA tube (9 ml) used for DNA extraction (whole blood)
d -	EDTA tube (9 ml) used for e.g. vitamins, antioxidants (plasma)
e -	EDTA tube (3 ml) used for HbA1c (whole blood)
f -	tube holder
g -	needle
h -	Plastic tubes (short-term storage - 20°C)
i -	Cryogenic vials (long-term storage -70°C)
j -	storage boxes (for whole blood tubes, plastic and cryogenic tubes and vials)
k -	Sheet of bar code labels (for blood and storage tubes and storage boxes)
l -	tourniquet, skin cleaner, pipettes, tips, skin tape, etc.

Items *a* to *g* must be supplied by the same national supplier and all tubes are evacuated. Items *h* to *j* should be compatible with the systems of both NHESL and EHES RL. Eg. the cryogenic tubes must be straight-walled in order to enable reading of the bar code.

10.5.2 Other equipment for handling, transfer and storage

- centrifuge, capable of 3,000g. If gel tubes are used, the centrifuge should have a swinging bucket rotor.
- racks for tubes
- special boxes for tube transfer and storage. The storage boxes should fit the freezer racks.

- set of (bar code) labels with identification codes or other method to mark the tubes (note that these should not be vulnerable to freezing)
- freezer (as required).

10.6 Processing, storage and transport of blood tubes

10.6.1 Processing

- Immediately after venipuncture, bar coded labels are placed on those blood collection tubes which have been successfully filled with blood.
- Centrifuge tubes at room temperature (20-25°C) for 10 min at 2000 g.
  - Plain serum gel tubes (a) are centrifuged within 30 - 60 min from venipuncture. Adherence to the time range and room temperature is necessary for complete clotting.
  - Plasma tubes (b,d) are centrifuged together with the plain serum tube within 60 min from venipuncture. Simultaneous centrifugation of both serum and plasma tubes ensures the identification and aliquoting of samples from the same subject.
  - Tubes c and e are NOT to be centrifuged
  - The caps should not be removed before centrifugation
- Immediately after centrifugation remove the caps. Place bar code labels on serum and plasma storage tubes. The labels should be fixed upright (see Figure 10.1).
- Transfer with pipette serum or plasma into the storage tubes according to a prefixed scheme, example in Figure 10.2. When using gel-containing tubes, it is convenient to pool the serum before pipetting into aliquots, as shown in Figure 10.2. Cap the tubes tight.

Note: g =(relative centrifugal force, RCF) is calculated from the formula:

$rpm = 1000 \times \sqrt{RCF/(11,17 \times r)}$, where

- r = radius, distance from tip of tube to center of rotor (cm),
- rpm=rotations per minute.

Figure 10.1 Labelling of storage tubes
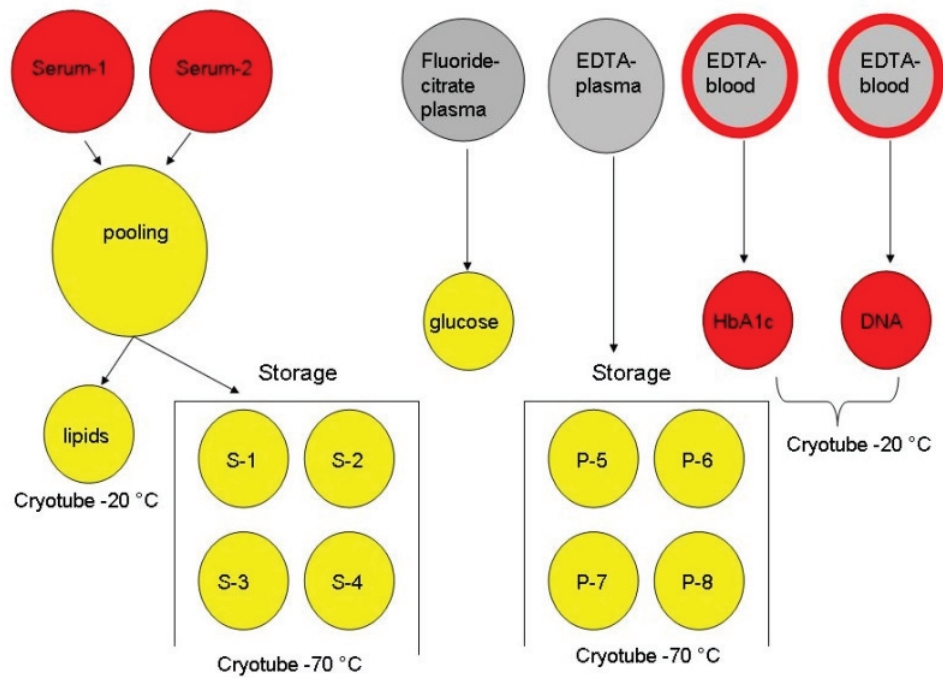
Figure 10.2 Example of blood processing

10.6.2 Storage of whole blood, serum and plasma tubes

Only a few aliquots of serum and plasma will be used for the core measurements. Samples frozen at -20°C should be analyzed within six months. These are typically reserved for the core measurements. For long term storage reserved for additional measurements and future use, the samples must be frozen at -70°C or less. Note that tubes intended for core measurements should not be discarded after analyses, but should be returned to the original storage temperature.

The storage boxes should be labelled with their appropriate bar code label BEFORE placing them in the freezer. Otherwise the labels will not stick. Place tubes upright in their designated boxes without delay and keep the boxes in a - 20°C freezer. An inventory of all stored specimen must be documented daily at the examination site.

10.6.3 Transport of specimen from the examination centre to the NHESL

The frozen samples should be sent in suitable batches during or at the end of the fieldwork. An inexpensive temperature check is to place a pre-frozen tube, half-filled with water, upside down in each batch. The boxes are transported to the NHESL in freeze or with an adequate amount of dry ice.

When the transportation is organized by a courier company, it should be arranged well in advance. A courier company will provide the service including necessary paper work for "door-to-door" transport.

Before sending the shipment, please inform the contact person of the receiving laboratory of date, courier and tracking details by email. The receiving laboratory should acknowledge the received shipments by e-mail or phone. The examination centre should keep a log book of all shipments, where also the acknowledgments are recorded.

10.7 Guidelines on laboratory performance

10.7.1 Performance of laboratories

Concerning the core measurements, the golden standards are values determined by the Centre for Disease Control (CDC, Atlanta). Data on the three following levels of accuracy (bias) performance ascertainment for core measurements will be monitored by the central EHES RL:

- Bias in EQA programmes of NHESLs related to the core measurement methods during the preceding year.
- Bias between NHESLs and EHES RL

- Bias between EHES RL and Centres for Disease Control (CDC, Atlanta).

This accumulated data will be reported annually.

10.7.2 Standardization of the laboratory analysis

1. When the EHES TL is in place, a pilot calibration between the NHESLs and EHES RL is carried out before the beginning of the survey. Therefore, the NHESL should contact the EHES RL at least 6 months prior to the planned starting date of the national HES. If the pilot calibration is satisfactory, proceed to step II. If it is not satisfactory, continue calibration pilots until results agree.

    The calibration consists of a series of reference samples having target values.

2. Depending on the analyte and number of survey participants, a random 5 or 10% of actual survey samples are transported and reanalyzed at the EHES RL.

10.7.3 Recommendation for analytical methods

No recommendation to use a specific method is given. However, only validated methods should be used, and the procedures should be documented. The documentation for each method used in the survey should be available at the NHESL and the EHES RL.

10.7.4 Quality Control

Data on precision of methods within and between days (series) must be kept with a computer-aided protocol. The goal and acceptable precision between days(series) for the core and additional analytes are shown in Table 10.2. The precision limits provide guidelines for the performance of the method. The limits are based in part on data from instrument manufacturers and experience of the EHES RL.

10.7.5   Accuracy (bias) and external quality assessment (EQA)

Data and documentation on accuracy of the methods are provided by participating regularly, more than once per a year in national or international EQA programmes, such as standardization procedures described in Section 10.7.2.

It is recommended to check and document the performance of all instruments (clinical chemistry analyzer, photometers, balances and pipettes, eg.) once a year. As with all components of a method, the

traceability of calibrators should be documented in order to ensure high quality.

The recommended goal and acceptable bias values for the core and additional analytes are shown in Table 10.2. The data on bias take into account the biological variation of the analyte. The values are modified from the reference: www.westgard.com/europe.htm.

Table 10.2. Recommended goals for bias and precision of methods

	Bias (%)		CV (%)	
Core analytes	**Goal**	**Acceptable**	**Goal**	**Acceptable**
Serum total choles-terol	3	5	2	3
Serum HDL- cho-lesterol	5	10	2	3
Plasma glucose	4	8	1	2
Additional ana-lytes				
Apo A-I	5	12	2	3
Apo B	5	12	3	5
Serum Tg	7	15	3	11
Blood HbA1$_C$	2	3	2	3

10.7.6 Corrective measures

If the bias of a method exceeds the acceptable value, corrective measures should be performed. Likely errors stem from erroneous calibrators, change of reagents (kits), malfunctioning instruments, wrong type of sample (serum-plasma-whole blood) and reagents not compatible with the instrument. An unexpected shift in bias may be observed, for example, when new calibrators are introduced. Also change of the method (technique) may provide remedy. According to good quality criteria of a laboratory, all NHESLs must document all changes done to methods or procedures and report promptly to the central EHES RL.

10.8 Collection of urine specimens

The major constituents of urine in decreasing order are urea, chloride, sodium, potassium and creatinine. The composition of urine varies strongly during the day due to physical activity, diet, exercise and rest. Many body constituents are affected by circadian variation, such as hormones and electrolytes. The mode of collection of urine is dictated by the analytes to be measured. In population health surveys commonly the following analytes have been measured: sodium, potassium, albumin, creatinine and iodine.

Usually urine specimens are collected over a predetermined time interval, such as 24 hours, morning urine after overnight fast or spot urine. For estimation of salt intake a 24 hour collection is required. The daily

intake of certain other dietary nutrients can also be estimated by this method. It is generally the recommended method of collection, but is cumbersome for the subject. However, this type of collection provides a way to check the completeness of collection by measuring the excretion of gender and age-specific creatinine which is fairly constant (Remer et al. 2002).

Morning urine has the advantage of being much more concentrated than a random spot urine specimen. It is also less prone to be influenced by diet. A spot urine sample can be obtained any time of the day. Its disadvantage is that it can be highly diluted and this cannot always be overcome by creatinine measurement. Nevertheless, both morning and spot urine accurately reflect soy intake (Franke et al. 2010). Urinary iodine from a spot sample is an established biomarker for its dietary intake (Lazarus 2015).

A 24 hour collection of urine requires careful instructions as to duration of collection, equipment and storage, less so for morning urinary collection. Spot urine is the least troublesome for subjects participating in health surveys and will thus result in a high compliance. Use of preservatives to stabilize urinary constituents depends on the analytes in question. Usually their function is to either lower or elevate the pH. Urine specimen may be sent to the laboratory by mail where the samples are stored preferably at – 70 °C. Loss of albumin is reported during extended storage at – 20 °C (Martin 2011), but no loss has been reported for creatinine (Remer et al. 2014). Instructions of the laboratory on preanalytical conditions and collection should be followed.

Urine can be used to monitor microbiological infections as well as environmental contaminants. Preanalytical precautions in microbiology include washing genitals before voiding and for e.g. environmental trace elements, use of metal-free storage containers.

10.9 Safety and laboratory quality procedures

Guidance on issues regarding safety procedures of laboratory personnel and laboratory quality assessment, eg. equipment are given in Part B, Section 5.5.

References

- Remer T, Neubert A, Maser-Gluth C. Anthropometry-based reference values for 24-h urinary creatinine excretion during growth and their use in endocrine and nutritional research. Am J Clin Nutr 2002;75:561-9.
- Franke AA, Hebshi SM, Pagano I, Kono N, Mack WJ, Hodis HN. Urine accurately reflects circulating isoflavonoids and ascertains compliance during soy intervention. Cancer Epidemiol Biomarkers Prev 2010;19:1775-83.
- Martin H. Laboratory measurement of urine albumin and urine total protein in screening for proteinuria in chronic kidney disease. Clin Biochem Rev 2011;32:97-102.
- Lazarus JH. The importance of iodine in public health. Environ Geochem Health 2015;37:605-18.

- Remer T, Montenegro-Bethancourt G, Shi L. Long-term urine biobanking: Storage stability of clinical chemical parameters under moderate freezing conditions without use of preservatives. Clin Biochem 2014;47:307-11.

11. Quality assurance

PART A

Kari Kuulasmaa[1], Päivikki Koponen[1], Hanna Tolonen[1]

[1] National Institute for Health and Welfare (THL), Helsinki, Finland

Quality assurance of a heath examination survey (HES) refers to the measures that are undertaken to ensure a good quality survey. Well planned quality assurance is essential in order to obtain high quality data, which will be comparable between countries and in particular over time, so that reliable long-term trends can be calculated from the data in the future. The basic components of quality assurance are:

- **Good overall management** of the survey.

- **Agreement on survey procedures** that ensure standard-ized stable measurements. These, together with all other requirements for the national survey management, should be described in the national manual of operations.

- **Training** of the survey personnel on using the standard procedures.

- **Piloting** the fieldwork phase, including also data manage-ment, transfer of biological samples and quality control.

- **Quality control**, which refers to measures taken to monitor the survey process, so that any problems can be detected at an early stage. The term 'quality control' also includes the action taken to correct the detected problems. In the ideal case, the problem will be detected early enough so that it can still be remedied. We will also use term **Quality as-sessment**, which refers to the monitoring and documenting the quality, but does not include the corrective action.

- **Evaluation of the achieved quality.** This includes also the documentation of the actual procedures used in the surveys, which may differ from those initially specified in the national Manual, as well as any corrective actions conducted on the bases of the evaluation. This step is necessary in order that the results of the survey can be interpreted correctly, taking into account the limitations of the survey quality. This step is also important for the documentation of the experiences, so that similar problems can be avoided in future surveys.

A separate section will be devoted to each of these components below.

Quality assurance should be seen as an integral component of all phases of a HES (Figure 11.1).
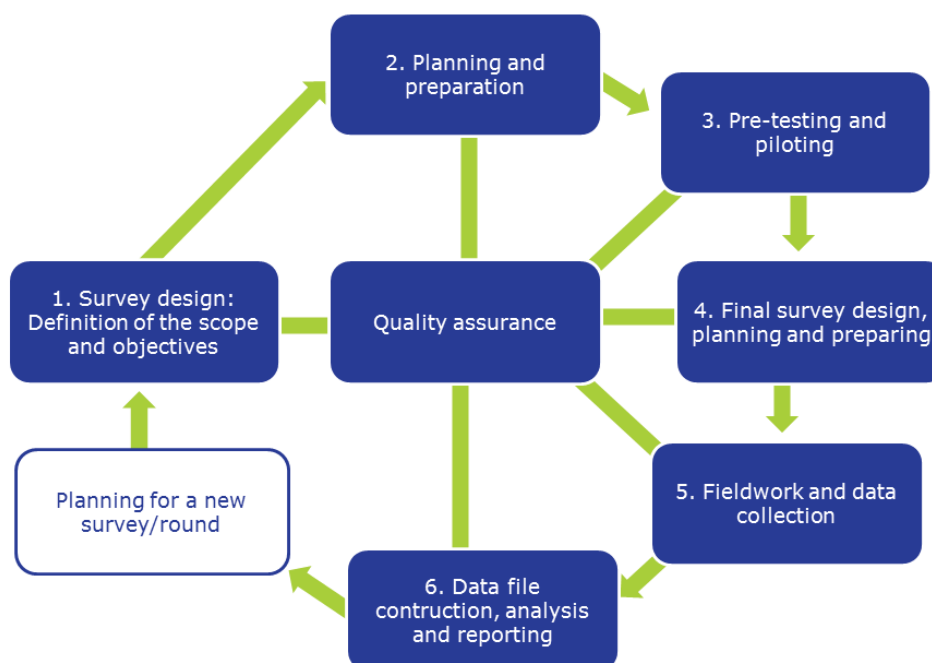
Figure 11.1. Role of quality assurance in a HES

11.1 Good overall management

Survey management is considered in Part A, Chapter 1. From the point of view of quality assurance, it is important that there is a well-defined survey organization. This includes a management structure, clearly defined responsibilities of the survey personnel and professional coordination.

A careful risk analysis helps the survey management to anticipate and prevent many problematic situations, which could otherwise have serious implications to the quality of the survey. Risk analysis, which is considered in Part A, Chapter 1 should be done in the planning stage of the survey and reviewed periodically during the survey.

11.2 Agreement on survey procedures

European guidelines and standards for the survey procedures are described in this EHES Manual. The national HES manual should describe the details of the procedures to be used in the national HES, also for issues where the EHES Manual can only give alternatives or general guidelines.

A template to facilitate the preparation of the national manual is available in the EHES website at http://www.ehes.info/. The national manual would usually be written in a local language. An English translation would be needed for European level evaluation, and its publication is encouraged as a reference of international publications of the survey results and as examples for other countries preparing their national manuals.

Whenever a country considers a procedure which deviates from the European recommendation, the issue should be discussed with the proposed EHES Reference Centre provided that it has been resourced for such a task. The decided deviations together with the justifications should be documented in the national HES manual. Countries are also encouraged to discuss such issues in the EHES extranet with other countries which may face a similar situation.

11.3 Training

Training is needed for acquiring the skills and motivation to follow the survey procedures. Each country is responsible for training the national survey personnel. Specific training is necessary in particular for the persons performing survey measurements in the field. The training and certification procedures for each measurement are described in Part B, Chapter 5 under the each measurement procedure.

If the field work takes more than few months, re-training sessions are usually needed. Measurement practices of the survey personnel often have a tendency to change over time. Re-training will reinforce the use of the standard procedures.

Recommendations for Europe wide and national training programmes are considered in Part A, Chapter 15 of the EHES Manual.

11.4 Piloting the HES

Piloting the HES and detailed evaluation of the pilot is crucial to ensure a successful data collection and field work phase of the HES.

Each country should carry out a pilot survey prior to the full-size HES. The purpose of the pilot survey is to evaluate the entire survey process and to obtain additional information for the planning of the actual survey. The sample size of the pilot HES should be estimated in such a way that it will lead to about 200 participants.

The aims and content of the pilot survey depend on the contents of the survey, previous experience and frequency of surveys. At least the following issues need to be considered (adapted from Primatesta et al. 2007, Biemer et al. 2003):

- Identifying need for further quality assurance activities, such as further specification of the recruitment, measure-

ment and training procedures. This facilitates planning the training of the survey personnel and finalizing the survey manuals.

- Getting feedback from the invited participants. This may concern the willingness to participate, recruitment process and information leaflets, informed consent and experiences on the measurements. The feedback is needed to develop different ways to motivate participation in the population.

- Recording timing and calculating average duration of interviews and examinations per participant. This is needed to estimate the duration of the full-size survey, the need for personnel resources (which has implications to budgeting) and potential burden to participants.

- Testing the use of equipment, computer programmes, data management, and the processing, transfer and storage of biological samples. This is needed to avoid problems in data and sample collection and management, and to assess the need for storage space, equipment and logistics.

- Identifying potential practical problems so that they can be avoided during the fieldwork. This may also give rise to refine practices and add further specifications to the national HES manual.

- Checking the data to identify needs for editing the questionnaires.

- Testing the analysis and reporting system to enhance rapid reporting of results. (This may not be feasible for all aspects of the reporting because of the small number of pilot participants.)

The survey 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 and readability,

- the choice of administration mode, and

- the respondents' burden (i.e. how long it takes to complete the questionnaire).

An optimal timing for the pilot survey is about six months before the full-size HES, so that there will be sufficiently time to evaluate the pilot and to make the necessary adjustments to procedures before the full-size survey. Any such adjustments to the procedures or instruments will still need to be tested prior to the full-size survey.

The countries should also consider the need for a small pretesting before the actual HES pilot. Such a pre-testing may be needed to test the computer programmes, measurement techniques and timing with a small number of volunteer participants.

11.5 Quality Control

The term "quality control" refers to the measures taken to monitor the survey process so that any problems can be detected at an early stage. Well planned and conducted quality control will save resources because it will minimize the time and resources needed for detecting and solving problems and for repeating tasks. It will also minimize the need to reject survey data because of loss or poor quality. Quality control is also needed to convince the users and reviewers (e.g. funders, publishers or results) of the survey data about the good quality.

11.5.1 Quality control of the planning of the HES

The main quality control activities of the planning stage are:

- to check that the plan for the preparatory phase of the survey covers all relevant aspects with sufficient detail (see Part A, Chapter 1); and
- to monitor the time schedule of the planning and preparatory phase.

One of the responsibilities of the proposed EHES Reference Centre (RC, see Part C, Section 6.2 of the EHES Manual) is to assist the national survey organizers by monitoring the progress of the planning and preparation of the national HESs and reviewing the national HES manuals. Accordingly, provided the EHES RC is resourced for this, each country should provide the EHES RC with:

- a schedule for planning the national HES as early as possible, and preferably one year prior to the beginning of the field work of the survey, and
- an English translation of the national manual preferably six months before the planned start of the survey.

The EHES RC will have to review and comment on the proposed schedule and the manual without delay, and in any case within three months after receiving them.

11.5.2 Organizing quality control for the survey procedures

Most of the quality control is carried out by the national survey team. This is called **internal quality control**. In addition, it is important to have the survey observed and assessed by an independent outside body. This is called **external quality assessment**.

11.5.2.1 Internal quality control

In principle, quality control is relevant for all phases where data are transferred from one form or place to another, and also for the re-

cruitment of participatns to the survey. Therefore, quality control is relevant for:

- recruitment of invitees to participate in the survey,
- the interview and measurement instruments and procedures,
- data and sample handling and transfer, and
- the data management.

Procedures for the internal quality control for each of these are specified in the respective Chapters of Part B. Persons responsible for internal quality control have to be assigned and documented.

The corrective action must always be thoughtful; to make sure that it really corrects the problem. A wrong correction may add a new component to the measurement bias. For measurements which involve a subjective component, such as interview or blood pressure measurement, the way of approaching the measurer needs to be planned carefully in order to prevent over correction. The best action can be for example a routine retraining of all personnel doing the measurements. It is often better to retrain the relevant group rather than to point out a single measurer.

The activities of internal quality control should be documented in a log book, together with any concerns detected and the action taken to correct problems. Examples of such log books are given in the Part B, Section 5.6.

The implementation of the internal quality control in each country should be described in the national manuals.

11.5.2.2 External quality assessment

External quality assessment is never a substitute for the internal quality control. It complements the internal quality control by providing an independent review of the performance, checking that the national standards are similar between the countries and over time, and overseeing that the internal quality control functions as planned.

The EHES RC (providing it is available as proposed) coordinates and carries out external quality assessment in EHES. Individual countries may use additional sources of external quality assessment, but they should keep the EHES RC informed of this.

The external quality assessment carried out or coordinated by the proposed EHES RC includes:

- Monitoring the progress of the planning and implementation of the national surveys.
- Review of the English versions of the national manuals.

- Site visits during the national surveys to assess the survey and quality control procedures.
- Assessment of the data obtained from the surveys. This will be described in more detail in subsection "Evaluation of the achieved quality" below.
- External laboratory quality assessment. This is described in more detail in Part A, Chapter 10.

11.6 Evaluation of the achieved quality

The evaluation and documentation of the achieved quality involves analytical investigation of:

- the actual survey procedures used,
- the data generated in the surveys as well as related documentation, such as any corrections of the data on the basis of the internal quality control and the documentation of routine data checks in the case of computer aided data collection, (see Part A, Chapter 8), and
- the data and information generated though external quality assessment.

The evaluation report is an essential prerequisite for the analysis and correct interpretation of the survey data (see Part C, Chapter 3 of the EHES Manual).

For the overall conduct of the survey and for each of the core measurements, the evaluation would be carried out by the proposed EHES RC, with the help of national survey organizers. If no such EHES RC will be resourced, the evaluation should be done nationally. For any additional survey component, such evaluation should be carried out nationally or, when possible, in collaboration with other countries which have also added such components to their survey.

As part of the planning of the survey, there is a need to ensure that the data collection in the survey includes all data needed for the evaluation. For the recruitment and the EHES core measurements, the required data items are described in Part B of the EHES Manual, under each measurement protocol. For additional measurements, countries should define the required information before they start their field work.

References

- Primatesta P, Allender S, Ciccarelli P, Doring A, Graff-Iversen S, Holub J, et al. Cardiovascular surveys: manual of operations. Eur J Cardiovasc Prev Rehabil. 2007 Dec;14 Suppl 3:S43-61.
- Biemer PP, Lyberg LE. Introduction to Survey Quality. Wiley, 2003

12. Data management

Ari Haukijärvi[1]

[1] National Institute for Health and Welfare (THL), Helsinki, Finland

A well-organized data management is an essential part of a health examination survey (HES). It ensures that:

1. the data will be available for analyses, and that the available data are

   - *complete*. No data collected from the survey subjects are lost.

   - *correct*. There are only justified differences between the values which were originally measured and the values in the final data storage.

   - *verifiable*. The relationship between the original data collected and the data in the final data storage can be described.

2. the data analysis will be

   - done using the correct data and other information which are relevant for the data analysis.

   - done without errors.

   - documented in such a way that the whole analysis or a part of it can be repeated later. If the documentation is not done properly, it may be difficult or impossible to reach the same results when similar analyses are repeated in other situations.

3. the confidentiality of the data is secure.

Point one above involves data collection, checking of data, error correction, data transfer from the field to the final storage (database), documentation and back-up of the data. Point two concerns analysis of the final data to obtain survey results. *Separation of the data management into these two stages is recommended.* If the survey data are not completed and the quality of the data are not documented before the data are analyzed, it is likely that the analysis will reveal problems in the data, many of which could have been detected earlier. This in turn

can result in much longer delays in the final analysis than if more care had been taken during the preparation of the data. Furthermore, use of unchecked and uncorrected data will lead to incorrect results. Well-planned data management facilitates good quality and easy and promt availability of the data for analysis and rapid reporting.

Therefore it will be necessary to create a detailed plan for the data management including all phases of the survey. Planning of the survey data management should be part of the general planning of the survey from the beginning. The following things need particular attention:

- detailed data flow in the survey;
- transfer, storage and security issues during each phase; and
- rules for data correction (data correction procedures).

12.1 Basic work and data flow

Below the following topics are considered from the point of view of organizing national HES data management:

- sample selection and recruitment;
- appointment scheduling;
- collecting the survey data;
- error checking, correction, and documentation of the data;
- transfer and storage of the data.

It is assumed that each country establishes a database for their survey data and maintains it locally in the country. A database should be prepared to store individual level data on the national HES measurements (including the questionnaire part), information on the quality of the data, as well as sampling data of each survey respondent. The national HES database serves as the local repository for the data used for evaluation of the surveys and for analyzing the survey results. An example of the work and data flow in the national HES is depicted below in Figure 12.1.
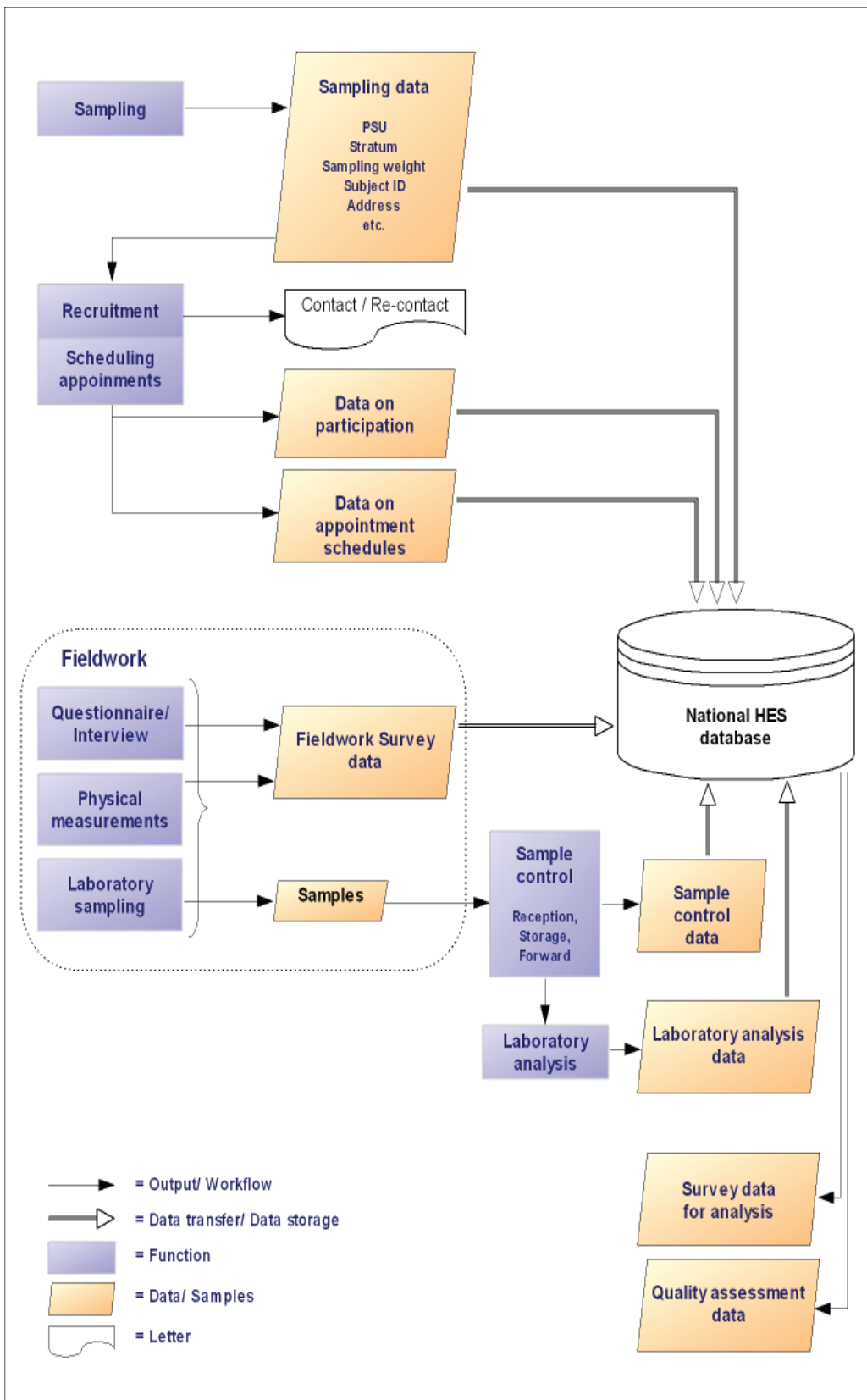
Figure 12.1 An example of HES work and data flow

There are several good methods to implement the stages of survey data management ranging from manual methods to computerized ones. In each phase modern information technology can be utilized. The choice

of the methods will depend on local facilities, existing practice and the expertise available.

12.2 Subject Identification

During the sampling, recruitment and examinations the persons are identified using the name, address, personal identification code, which is available in many countries, and possible other information which make it possible to contact and examine the selected person. After the sampling, also another unique Subject Identification code should be assigned. It must not allow the identification of the real person. The Subject Identification code should show up in all data records. After the fieldwork, the person identification which connects the records to the real person should be removed from the measurement data records. It should be stored separately and be accessible only to named authorized persons. Usually the Subject Identification code is the same as the SERIAL NUMBER, which is defined below.

When data are transferred to the EHES Reference Centre, the subject records will be identified using three levels of codes:

- **COUNTRY:** This identifies the country of the survey, using the EU-coding (see http://publications.europa.eu/code/en/en-370100.htm). It is the same as the two-letter ISO 3166-1 alpha-2 code with two exceptions: Greece is EL (not GR) and United Kingdom is UK (not GB).

- **SURVEY:** This two digit number code identifies different EHES-surveys in the country. It is assigned by the national survey organizers but shoud be confirmed by EHES Reference Centre before applying.

- **SERIAL NUMBER:** This number identifies the different persons selected to the sample. It is assigned by the national survey organizers, and is subject to the following principles:

  - SERIAL NUMBER is given to everybody selected to the sample (i.e. not only for example those eventually found eligible or to the survey participants.)

  - It is unique within the survey. Only one person selected to the survey can have the same SERIAL NUMBER. However, surveys in different countries or different surveys in the same country (identified by different SURVEY codes) can use same SERIAL NUMBERS. When addresses or households are used as sampling units, the part of the serial number identifying the subject may be completed only at the stage when the household is visited.

  - Because errors in the SERIAL NUMBER usually lead to a loss of the record, it is strongly recommended that the SERIAL NUMBER includes a check digit (or check digits). An example of a convenient single character numeric check digit detecting all one digit errors and all transpositions of adjacent digits has been described by Gumm (1986).

- The serial number should not include information which makes it possible to identify the person in the population. For example, person identification codes available in many countries should not be used as serial number.
- When data are transferred to EHES Reference Centre the maximum length of the SERIAL NUMBER is 12 characters.

12.3 Data sources

12.3.1 Sample selection, recruitment and appointment scheduling

12.3.1.1 Sample selection

The first major data management issue relates to sample selection and recruitment (see Part A, Chapters 2, 3 and 13). As a minimum, the following information is to be recorded for every subject selected to the sample:

- *The Subject Identification (e.g. the SERIAL NUMBER,* as specified in Part A, Section 12.2) should be given to everybody selected to the sample and used to identify the subject throughout the survey and data management.
- *Sampling information*, as specified in Part A, Section 3.8.
- Address information, such as the person's name, address and any other information will be needed to contact the person.
- Additional information, such as sex and age are also often available from the sampling frame.

The *subjects selected to the sample form the basis for the control of the data completeness* through the data management process. The survey history of every subject should be verifiable from the final database.

12.3.1.2 Recruitment

At the recruitment stage, eligibility status and attempts made to contact each subject need to be monitored, and the eventual success of the recruitment should be recorded. Also the contact information may need to be updated. For each person invited to participate in the HES, it is necessary to keep a record of the following information:

- Eligibility status;
- The timing, number and type of contact attempts (home visit, telephone call etc.), and who made the contact attempt;
- Contact status;
- Participation status;

- The reason for non-participation. For the subjects who did not participate in the survey examinations, the reason should be recorded using the classification listed in Part A, Section 13.3.

12.3.1.3 Appointment scheduling

Organizing the appointment schedule is necessary for a successful fieldwork examination and is closely linked to the recruitment phase. For example in a case when an appointment was fixed but the subject is nevertheless unable to participate in the survey, this changes not only the appoinment schedule, but also the participation status (or even eligibility status) in the recruitment data.
Here at least the following information should be logged:

- The subject identification (see Part A, Section 12.2).
- The contact information (name, address, phone number).
- The appointment information (time, place).
- The recruitment information (the record of participation).

Regarding the future surveys it is also important to keep a record of changed appointment times and the situation where an invitee tried to change the given appointment time but it was not possible to find feasible time for him/her.

There are several commercial applications for scheduling (e.g. patient scheduling software used by hospitals and clinics). Some of them are web-based scheduling services, while others are standalone client software. The usefulness of this software depends on how they can be customized to manage the necessary data. One possibility is to build a dedicated HES database application for the survey project to serve both the recruitment data and the appointment scheduling. The application should preferably interact with the national HES survey database. Change of experience between countries will be useful when planning or selecting an application for this purpose.

12.3.2 Survey data

12.3.2.1 Data sources

Recording the survey measurements and getting data from different examination sites to the common database are essential parts of the national HES data management. These include:

- signed informed consent forms;
- completion of self-administered questionnaires (if self-administered questionnaires are used);
- interview;
- web questionnaires;

- recording the values of physical measurements either manually or directly from the measurement device;
- biological sampling, processing and transfer to the laboratory;
- recording the laboratory results;
- transfer of paper forms and/or electronic records from the examination site or laboratory to the survey data centre;
- getting the data into electronic format and to the common database.

Three main challenges for the data management during these steps are to ensure that

1. no errors are made in recording the results;
2. the data records are complete;
3. no records are lost or different persons' records are not mixed up.

Errors and incompleteness of the records can be prevented by good design of the record forms and by routine checking of the forms and the data. The earlier the errors will be detected, the easier their correction is. When feasible, detection of errors should be done when the subject is still in the interview or at the examination site.

Relevant data which were not obtained from the subject should not be left blank, but a specific code for missing data should be used. Subsequently, the incompleteness of the data can be detected as blanks in the data forms.

To prevent the loss of records, it is important that the subject identification becomes correctly recorded at all stages and the Subject Identification code will be used. If feasible, laboratory samples (biological samples, storage tubes and storage boxes) should be labelled with bar codes with a reference to the correct Subject Identification code.

All steps where data are transferred from one form to another or from one place to another require specific attention when data management of the survey is being planned.

12.3.2.2 Forms of data collection

The procedure of collecting the survey data will be different according to whether the data are collected directly to computers or on paper forms.

- A computer-assisted data collection has the advantage of reducing the number of manual data transfers and facilitates extensive data checking at an early stage. However, such a system should be used only if it has been tested in the field and found reliable. Otherwise there will be an increased risk for losing records or delaying the examination schedule due

to breakdown of the system. Paper forms should always be on hand as back up in case of power failure or other problems with computer devices.

- The use of paper forms has proven to be reliable, but they have the problem that on-site data checking is more difficult. If paper forms are used, the typing of the data into electronic format needs to be done carefully. In this case the traditional double typing method by different persons is worth considering. Also optical character recognition (OCR) can be used to convert scanned images of handwritten or printed text into electronic files. This again sets up challenges for the design of the forms and data error checking, i.e. validating the OCR converted data.

Table 12.1 Computer-assisted interviewing methods (see also Part A, Chapter 8)

Method	Description
Computer-assisted personal interview (CAPI)	Interviewers meet respondents and conduct face-to-face interviews using a computer. There may be an online connection to an external database from the computer or the data are sent to a central computer after the interview (either via Internet or by sending data disks by mail).
Computer-assisted telephone interview (CATI)	The interviewer sits at a computer and asks the questions appearing on the screen. The respondents are on the telephone. The respondent's answer is typed by the interviewer. Supervisors are present for quality control and to assist with specific problems.
Computer-assisted self-interview (CASI)	In a computer-assisted self-interview or self-administered web-survey the respondents themselves read the questions on the screen and enter the answers. There is no interviewer; the interviewing programme guides the respondent through the questionnaire. This procedure can appear also as part of a computer-assisted personal interview session where the interviewer hands over the computer to the respondent for a short period, but remains available for instructions and assistance. This is similar to the procedure used in traditional face-to-face interviews where an interviewer might give the respondent a paper questionnaire containing sensitive questions.

The computer-assisted data collection may include both the interview and the measurement phase. Computer-assisted interviewing methods are described in Table 12.1. In computer-assisted data collection automatic built-in checks for responses become possible, and data entry by trained fieldworkers reduces errors. A computer-assisted interviewing system and survey processing tool can include features to define questionnaires, data validity and range checks, conditional error handling, etc. which facilitate both the questionnaire design and data entry, and can help to prevent printing errors.

Storing the final data as soon as possible after the measurement will make it possible to have only one recording round for the data, enhance data security, and ensure that the data will not be forgotten or lost.

12.3.2.3 Preparation of the fieldwork

The preparation of the fieldwork phase includes:

- Planning of the data management and data transfer system for the fieldwork (i.e. computers, network, software, and other equipment needed) and testing these
- Arranging the training, responsibilities and support of fieldwork teams

Computer equipment and network may need to be considered. Here things that require particular attention are data security, data transfers, and back-up of the data to avoid any losses of data in case of a system flaw. Relevant questions and issues are:

- Shall a private LAN/WLAN be established for each fieldwork team?
- What computer equipment will be needed?
- What software will be needed?
- Planning software update procedures during the fieldwork
- Planning data back-up equipment and procedures
- Planning the transfer of the data
- Data security: locking of computers, usernames and passwords to access the computers, encrypting data on computers, how to store paper forms on the field, etc.

Regarding the software, all workstation computers are recommended to be identical with each other which make them easy to replace if broken. Server computers (if any) in the fieldwork can be designed to be stand-alone, independent portable computers, easy to relocate or replace in case that one of them is broken. A backup mirror can be implemented between servers, in which case the data on the first server are copied to the second one.

The equipment and data transfer and storage systems should be tested thoroughly prior to the training of the fieldwork personnel. Time needs to be reserved both for testing and analysing the test results. A well-defined responsibility in recording the data is important.

12.4 Error checking, correction and documentation of the data

After the data collection, the data should be checked as soon as possible for

1. strange *values*, i.e. for values which have not been defined, and also for values which are possible but rare;
2. *consistency* between the values of different data items;
3. *completeness*, i.e. that all data items have been recorded and no records have been missed.

A visual checking of the key items can be done at the interview or examination site even if paper forms are used. An extensive checking should take place as soon as the data have been computerized. When potential errors are detected, they should be investigated for correctness, and corrected only if it is found that they really are errors. It is advisable to authorize only those who have made the errors to correct them, since they are usually in the best position to tell if there really is an error, and are often the only ones who know the correct value. Each error and its possible correction should be documented.

The frequency of errors, which were not possible to remedy should be documented. The same concerns the results of the quality control during the data collection, any deviations from the survey protocol, and any other information which may be relevant for the interpretation of the results. Knowledge of these issues is essential for those who analyze the data and interpret the results.

Each data transfer and import into the central national HES database should include at least a routine check for each data variable.

Examples of routine error checking criteria can be found in EHES Manual Part C. Documentation of the quality of the data in a multi-national setting can be found e.g. in the WHO MONICA quality assessment reports, which are available at http://www.ktl.fi/publications/monica.

12.5 Transfer and storage of the data

12.5.1 Data transfer and interface to import the data

Data transfer and import into the national HES database depends on whether the data are collected using computer-assistance or manually by using paper forms. When data or samples are transferred from one place to another, it is important that the data transfer is logged properly. All data transfers should be traceable whether they are computerized or manual. The recipient of the data or samples should be able to check that he or she has received exactly the same records which were sent. The person sending the data should make sure that everything

was received. The transfer of the data into the central data storage, the national HES database, should be done regularly and via a secure data transfer medium.

Ways to transfer computerized data are via Internet through a secure connection and data encryption or by storing data on a dedicated medium on which data files are delivered to the HES coordinating centre by a secure mail. Such media may be optical disks, SSD/Flash disks or USB memory sticks, but these require special concentration on the security of the transferred data.

- Web application - a dedicated web software interacting with the central HES database, file server or other database from which the data are further transferred into the central HES-database. This kind of an application typically functions on web browser.
- Direct import from data files. The data delivered to the HES co-ordinating centre are imported in the central HES database using standalone client application on dedicated computers.

An example of a possible HES database system architecture is depicted below in Figure 12.2.
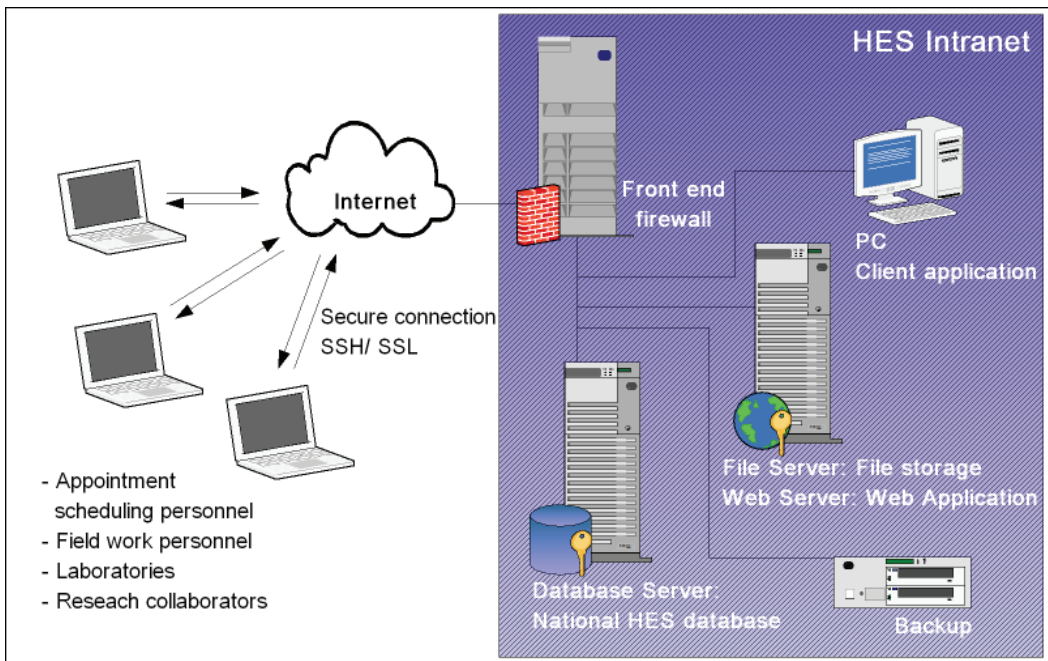
Figure 12.2 An example of HES system architecture

12.5.2 Data security

The core principles of information security are data confidentiality, integrity and availability regardless of the form the data may take: electronic, printed, or other forms. (Table 12.2)

Table 12.2. The principles of data security

Security principle	Description
Confidentiality	Ensuring that information is accessible only to those authorized to have access. This is necessary for maintaining the privacy of the people whose personal information the system holds and to prevent the disclosure of information to unauthorized individuals or systems.
Integrity	Safeguarding the accuracy and completeness of information and processing methods from unintentional, unauthorized, or accidental changes. Maintaining data integrity is essential to the privacy, security, and reliability of the data. Data integrity can be compromised by malicious users, hackers, software errors, computer virus infections, hardware component failures, and by human error in entering or transferring data. Therefore, the access to protected information should be done only through proper identification and authentication of the users.
Availability (the degree to which the system is operable)	Ensuring that authorized users do have access to information when required, i.e. the information is available. This can be accomplished utilizing data back-up plans and continuity/recovery plans.

To ensure data confidentiality and integrity it is necessary to use technical controls - e.g. passwords, network firewalls, access control lists, and/or data encryption - to monitor and control access to the computing systems and collected data. Some standards for these are described below in Part A, Section 12.6.3. For example during the fieldwork it is necessary to protect data on local computers' hard drive, or when transferring data over a network.

It is essential that the information connecting the survey data to the personal identification of the subject will be available only to persons who have authorized access to such data. Only authorized persons should have access to the data and all of them must understand the importance of the confidentiality of the data. After data collection, the information from which a person can be identified and the code connecting this information to the subject identification of the survey records, should be stored separately from the survey data, and maintained e.g. on an encrypted hard drive. Normally only few people need access to the person identification information, whereas the rest of the survey data will need to be accessed by all who analyze the data. Specific precautions should be defined for the handling and storage of paper forms in the examination site and elsewhere when these are used.

Regarding the administrative controls, approved written policies and guidelines on data transfer and confidentiality, see Part A, Chapter 4.

12.5.3 Data back-up

All data in electronic format should be backed up routinely for accidental breaks of the storage devices, failures in data transfer and unintentional deletion of data files. Especially during the fieldwork it will be important to back-up the data on local computers' hard-drives against accidental losses. Common situations where important data have been lost, although some back-up was in place, are:

- Loss of data during data collection or data processing because of absence of back-up at these early stages. Accidental loss of the back-up data together with the original data since the two were stored together, or the broken device or system which destroyed the original data was used to open the back-up data. The complete back-up data had already been replaced by the incomplete data before the loss of data had been detected. This may happen if the system for long-term back-up is incomplete. There was a back-up, but there were insufficient documentation on its location or on the procedures needed to retrieve the data from the back-up. This problem could arise due to unforeseen changes in personnel.

Today there are more backup options than ever before. Technically several storage media can serve as back-ups:

- External/clone hard drive
- Optical disks (CD, DVD, Blu-Ray disk)
- Another computer dedicated to back-up purpose
- Magnetic tapes (LTO-7 provides 6.0 TB in a cartridge)

Back-ups are needed not only for data in electronic format, but also for important paper documents, such as log books of the survey examinations.

12.6 Recommended standards, techniques and tools

12.6.1 Database

The national HES database can be structured in different ways depending on the available facilities. The database should be created using a well-established database management platform, designed for scalability and extensibility.

Recommended standards for database construction are:

- Well-established relational database management systems, such as PostgreSQL (http://www.postgresql.org), Oracle (http://www.oracle.com), MySQL (http://www.mysql.com)

or Microsoft SQLserver (http://www.microsoft.com/sqlverver).

- Language to implement database structures and logic: ANSI SQL (http://en.wikipedia.org/wiki/SQL#Standardization).
- Standard database connection interfaces, such as ODBC (http://en.wikipedia.org/wiki/ODBC) and JDBC (http://en.wikipedia.org/wiki/JDBC).

12.6.2 Development tools, use of statistical software and XML

Several programming languages and development environments, as well as dynamic web content technologies exist to be used to implement an appropriate application logic and user interface for the national HES database ranging from JavaScipt (http://www.oracle.com/us/technologies/java/index.html), Java (http://www.oracle.com/us/technologies/java/overview/index.html) or Scale to .NET (http://www.microsoft.com/net) solutions. The choice will depend on the local facilities, existing practice and the expertise available.

It is recommended to use the XML-based general standards when implementing dynamic web content output and/or interchanging data over the Internet. In the survey data analysis it is recommended to produce analysis and reports by a well-established statistical software, such as R (http://www.r-project.org/) or SAS (http://www.sas.com/). SDMX (Statistical Data and Metadata eXchange, http://sdmx.org/) technical standards provide technical specifications for the exchange of statistical data and metadata based on a common information model.

12.6.3 Data encryption

The system should enforce security through data access control and auditing:

- *Access control* to restrict access to the data.
- *Auditing* to log the actions and changes which have been performed, when and by whom.

Data encryption may be necessary, for example to protect data on local computers' hard-drive during the fieldwork, or when transferring the data over network.

- To secure and encrypt data connection e.g. the following techniques can be used:
  - SSH (Secure Shell network protocol, http://en.wikipedia.org/wiki/Secure_Shell)
  - SSL (Secure Sockets Layer, http://en.wikipedia.org/wiki/Secure_Sockets_Layer) or TLS (Transport Layer Security, http://en.wikipedia.org/wiki/Secure_Sockets_

Layer,   http://en.wikipedia.org/wiki/Advanced_Encryp-
tion_Standard)

- AES (Advanced Encryption Standard)
- Blowfish (Fast block cipher, http://en.wikipedia.org/wiki/
Blowfish_(cipher))
- A web application can be built on an information server with
SSL/ TLS support to ensure encrypted connections to the
server. Recommended protocols for data transfer between
client and web server are HTTP/ HTTPS (by SSL/ TLS) and
FTP through SSH.

12.7 Local data management and the EHES Reference Centre

The collection of pseudonymous individual level data from each coun-
try to a centralized database at the EHES Reference Centre is neces-
sary for data quality assessment and for assessing the success of the
standardization and documentation of country-specific characteristics
of the data. This is described in Part C of the EHES Manual.

References

- Gumm HP. Encoding of numbers to detect typing errors. Int J Appl Engng
Ed. 1986;2:61-65.

13. Recruitment of participants

PART A

Sanna Ahonen[1], Päivikki Koponen[1], Hanna Tolonen[1], Kari Kuulasmaa[1]

[1] National Institute for Health and Welfare (THL), Helsinki, Finland

This chapter provides general guidelines for recruitment process, recruitment methods, definition of participation rate and non-participant data. The strategy and methods for recruitment have to be determined by each country based on national and regional feasibility and legislation, the survey budget, and cultural norms.

A high participation rate is fundamental to the reliability and validity of the survey. The participation rate depends directly on the success of recruitment. Proper recruitment is also necessary for the HES to be ethically acceptable. Description of the recruitment process is a key element in the research proposal reviewed by the ethical committees. The recruitment process needs to be carefully prepared and piloted.

13.1 Recruitment process

The purpose of the recruitment process is to ensure as high participation rate as possible and as representative group of participants as possible. The recruitment process includes all stages, where persons selected to the sample are contacted to provide information and to make appointments for examination visits. The recruitment process varies between countries and should be planned according to what is the most feasible and culturally acceptable way in each country. In some countries, there are also legal restrictions regarding to contacting potential participants.

13.1.1 First contact attempt

It is important to obtain as high participation rate as possible already with the first contact attempt. A successful first contact, without the need for additional attempts, saves costs. The first contact attempt can be made for example by an invitation letter and information leaflet and then be followed by a phone call in order to schedule an appointment. The written material that is used in recruitment can be divided as follows:

- **Information leaflet:** A leaflet that contains key information on the survey in a concise form, targeted to selected persons or also targeted more widely to stakeholders. It should be visually attractive, but easily distinguishable from advertising materials. It typically addresses:

  - Objectives of the survey

  - Brief description of the measurements

  - Importance of the survey for improving public health

  - Importance of participation

  - Benefits for the participant

  - Information on receiving personal results and how the survey results will be reported

  - Information on partners and financial support

  - Name and signature of the leader of the survey or some important person from the community related to the survey such as the minister of health, the head of health district etc. At least for some population groups, this may help to point out the importance of the survey

  - Strict confidentiality of survey data

  - Website address for more information and possibly for scheduling appointment

  - Contact information (toll-free number for more information, e-mail)

  - See an example of the information leaflet (at the end of this Chapter).

- **Invitation letter:** This is a personal invitation to participate in the survey. The invitation letter can be short, if other relevant information is given in an attached information leaflet.

  - Information on how the person was selected.

  - Pre-scheduled appointment time (with contact information for rescheduling) or instructions how to schedule the appointment.

  - Name and signature of the survey leader (or other important/respected person)

  - See an example of the invitation letter without pre-scheduled appointment time (at the end of this Chapter) and with pre-scheduled appointment time (at the end of this Chapter).

It is also possible to combine the information leaflet and the invitation letter.

- **Instructions to participant:** Instructions on practicalities regarding the participation.

- Includes information on examination (fasting, instruction to the examination site etc.) (see Part B, Section 2.1.).
- These instructions may also be included in the invitation letter.

- **Information sheet**
  - Provides the necessary information to participant before obtaining informed consent (see Part A, Chapter 4.)

Invitees' response to the first contact attempt highly depends on the contents and the format of the written materials. The materials should be informative, but also easy to understand, even by participants with a slight linguistic or cognitive impairment. The format, length and wording of the invitation could be modified according to the age of the participants. The material should be translated into all relevant languages, especially when the country has more than one official language and/or there are other major ethnic minorities without sufficient skills in the official languages. If, for example for financial reasons, it is not possible to translate all the material into several languages, it should be considered if at least the invitation letter is provided in several languages.

13.1.2 Re-contact attempts

Figure 13.1 shows different responses to contact attempts. Regardless of how successful the first contact attempt is, at least 1-3 re-contacts should be made if feasible and not restricted by national legislation. In all types of surveys, incorrect addresses and difficulties to obtain telephone numbers are well known problems. Many of the persons who do not show up after the first contact attempt, probably simply did not receive or open the invitation letter. If feasible, the envelope of the invitation letter could include a note for the post office informing that for recipients who have moved, the letter should be returned with information on the new address rather than being forwarded. For re-contact attempts, the accuracy and the recentness of the contact information must be checked, if possible.

The re-contacts may consist of a letter (with or without the questionnaire), phone calls or home visits, depending on the cultural acceptability and available resources. The availability of numbers for mobile phones and, whereas fixed telephone-lines varies by country and age-group. A personal approach is usually more effective than a second letter of invitation and allows the scheduling of the appointment to be "tailored". If a second letter of invitation is sent, it should be modified (e.g. introduction, signature, and format) compared to the first letter of invitation. The hours in which the measurements are taken can be made more flexible (early mornings, evenings, weekends, drop-in visits). Home measurements or visits to institutions (e.g. hospitals, nursing homes, prisons) may be offered if the person is unable (e.g. health condition) to participate otherwise. Reimbursements, incentives or small gifts additional to those used in the first invitation should be considered, if they are considered ethically acceptable and feasible. If a selected person refuses to participate in the survey, it should be

respected and recruitment attempts should end at that point. However a short non-participant questionnaire (see example at the of the this Chapter) may be offered to those who refuse.

Substitution of a non-contact with, for example, a neighbour or a person with similar characteristics (e.g. sex and age), is not acceptable (see also Part A, Chapter 3). Obtaining information from proxies for the key interview components of the HES is generally not acceptable (e.g., information on health issues provided by the spouse for a person working abroad). However, the non-participant questionnaire, may be answered by a proxy if the selected person cannot be reached or is otherwise incapable to answer. In addition parts of the interview can be answered by a proxy if the selected person is unable to answer due to e.g. limited cognitive functions (see Part A, Chapter 8).
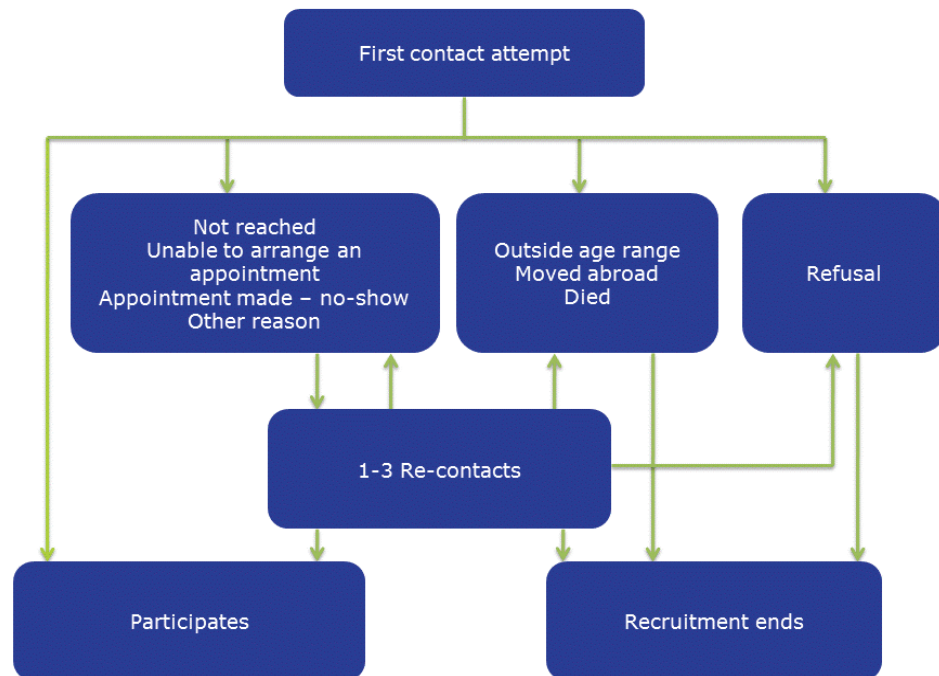
13.1.3 Refusal conversion

Refusal conversion refers to the situation when an invitee has been reluctant to participate but has not explicitly refused. In these cases, further contact attempts can be made, trying to motivate the invitee to participate. Refusal conversion often requires excellent communication and negotiation skills from the survey personnel. If an invitee tells that he/she refuses due to lack of time or is concerned about the confidentiality of the survey results, well trained survey personnel have tools to solve these. They can offer more flexible times for the examination and provide enough information about legal obligations relating to data security and ways how the data is handled in the survey to convince the invitee of the confidentiality of his/her results.

13.2 Participation rate

Participation rates should be calculated separately for the interview/questionnaire information and for examinations whenever feasible.

Following definitions are for the participation rates based on individual level samples, not household samples. Individual level and household level participation rates are not directly comparable.

13.2.1 Definition

Figure 13.2 shows the classification of the original survey sample. The definitions are:

- **Eligible:** A person is coded as eligible, if she/he belongs to the target population (see Part A, Chapter 2.).
  - **Participant:** An eligible person is coded as participant if she/he has at least one valid examination measurement, such as height and weight, in addition to some questionnaire items.
  - **Non-participant** is a person, who refused or otherwise did not participate after the invitation was assumed to have been received as some other contact was established. (Tolonen 2005, Wolf 2005)

- **Not eligible:** A person selected to the sample is coded as not eligible if she/he does not belong to the target population. This includes over-coverage of the sampling frame (i.e. persons who are in the sampling frame although they do not belong to the target population, e.g. not within the age limits) and persons who died or moved out of the primarily sampling unit (PSU) prior to the scheduled examination. The reason for being not eligible should be recorded. Persons who were temporarily absent during the survey period because of work, studies, tourism, hospitalization, or for other reasons are part of the target population and therefore eligible.

- **Unresolved:** There may be persons whose eligibility status cannot be resolved. In a typical case,
  1. the invitation letter was returned to the survey administration indicating that there is no such person in the address; AND
  2. other contacts were not possible or not successful; AND
  3. no information was available to assess the eligibility status.

Although it may be likely that the person does not belong to the target population, there is no certainty about this. The number of unresolved persons is usually small, but may be substantial in some countries, where good sampling frames are not available. or for some popula-

tion groups who often have un-accurate or outdated information in the sampling frame (e.g. students, institutionalised persons or immigrants without permanent personal address).
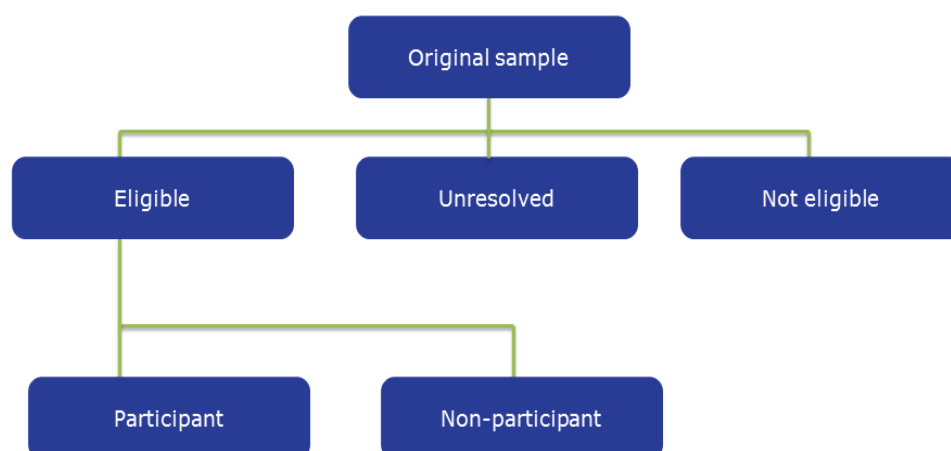
Figure 13.2. Classification of the original survey sample

The formula to calculate participation rate (PR) and its fractions co-operation rate and contact rate are shown in Table 13.2.

Table 13.2 Calculating participation rates

Participation rate = (number of participants) / (number of eligible AND unresolved)
Co-operation rate = (number of participants) / (number of eligible)
Contact rate = (number of eligible) / (number of eligible AND unresolved)
Note that Participation rate = Contact rate x Co-operation rate

13.2.2 Target participation rate

Recruitment efforts should be geared towards obtaining the highest possible participation rate so that the sample will represent the target population. The target rate of participation should be at least 65%, but preferably higher (see also calculation of sample size in Part A, Chapter 2). It should be noted that additional to high participation rate it is also important to understand population profiles of the survey non-participants. If there is indication that participants differ from non-participants on important variables such as health factors, the rate should be closer to 80% or over (Tolonen 2005). It is known that non-participants are more often young, men and from lower socio-economic class when compared to participants (Shahar 1996, Jackson 1996, Tolonen 2005, Eaker 1998). Non-respondents have also worse health profile, more psychological disorders (van der Akker 1998, Shakar 1996), are more often smokers (Tolonen 2005, Barchielli 2002, Macera 1990) and have higher total and cause-specific mortality than participants (Cohen 2002, Hara 2002, Harald 2007, Jousilahti 2005).

Previous HESs have shown great variations between participation rates among European countries. Only a few surveys have reached participation rates of 65% or higher during the last few years (HIS/HES database, Mindell 2015, Tolonen 2015). This is why special attention should be given to developing actions which may help to obtain the highest possible participation rates.

13.2.3 Ways to increase participation

13.2.3.1 Selection and training of personnel

Competent and motivated survey personnel play an important role during the recruitment process. The selection of fieldwork personnel has to be based on general requirements and competences needed to carry out the fieldwork tasks, as stated in Part A, Chapter 9. Good social skills, especially good communication skills, are prerequisites when selecting survey personnel. . If the target group includes large minority groups, their special needs relating to culture and language should also be considered when selecting survey personnel. (Corbin-Smith 2007)

After selecting competent personnel, sufficient training must be provided (See Part A, Chapter 15.). It is important, that the personnel responsible for recruitment as well as all fieldwork staff are familiar especially with the following issues:

- Understand the importance of a high participation rate to survey quality.
- Ways/ actions how to motivate participation
    - Know the correct answers to frequently asked questions about the survey and about participation to the survey.
    - Know what options can be offered in case of difficulties in scheduling a visit (e.g. weekend and evening hours, drop in or home visits).
    - Know all incentives and reimbursements available (e.g. if travel expences can be reimbursed for those who are otherwise not able to participate).

If there is a need to motivate the personnel responsible for recruitment, a bonus or other gifts may be considered, if feasible, for obtaining high participation rates, especially in districts or age groups where participation is expected to be lower.

13.2.3.2 Factors affecting participation rate

There are several actions that can be used to reach the target participation rate. In addition to the importance of the survey for serving public health and research, potential participants are also interested in personal benefits. For some participants a possibility to receive information on their own health status and risks may be an important reason for participation. Therefore, inclusion of additional examinations

which offer more information to the participants on their health status should be considered.

Incentives can also be used to motivate the invitees to participate. There is some evidence that unconditional incentives, incentives which are given prior to the participation are more effective than conditional incentives which are given only after participation. (Edwards 2009) As monetary incentives aren't often accepted by Ethical Boards, different tokens of appreciation may be considered (such as gift cards) rather than payments.

Table 13.1 Factors that may affect participation rates

Factor	Possible effects on participation rate
Pre-notification	Pre-notification either by mail or SMS prior to invitation to participate in the survey usually raises the participation rate (Phillips 2002, Spry 1989, Tolonen 2014).
Phone call	Phone contact is an effective way of increasing participation rate (Heistaro 2008, Lundqvist 2016).
Multiple contacts	Multiple contacts significantly increase participation rates (Porter 2004).
Flexibility in scheduling appointment	Offering evening and weekend times, drop in visits and different locations for measurements increases participation especially among busy people (Heistaro 2008).
Relevance and importance	Survey relevance and importance to the survey recipient is an important factor when designing surveys and key messages. Highly relevant surveys raise the participation rates. (Porter 2004, Phillips 2002)
Personal fulfilment	Feeling valued and appreciated increases the willingness to participate (Phillips 2002). Signature or introduction in the invitation letter written by a respected person may increase the feeling of being valued.
Statements of confidentiality	Loss of privacy when providing biologic specimens can be a major concern affecting participation rate. This is why it is important to explain confidentiality issues to the participants (Samanic 2003).
Requests for help	People with personal tendency to altruism tend to follow a norm of social responsibility and may be more willing to take part in the survey, if a phrase "it would really help us..." is used in the invitation (Porter 2004, Sinicrope 2009).
Sponsorship	Surveys sponsored by academics or governmental organizations have higher participation rates in general than surveys sponsored by commercial organizations (Porter 2004).
Mass media campaigns	Raising public awareness about the survey: the importance in national, community and individual levels.
Home visits	Home visits raise the participation rate if a person is unable (e.g. difficulties in functional capacity) or unwilling to participate otherwise (Heistaro 2008, Lundqvist 2016), or when people prefer or are used to home visits in their health services.
Domestic vs international use of research samples	Participants may be more willing to allow samples to be used for domestic rather than international studies (Tupasela 2009).
Several languages	Using several languages helps in recruiting ethnic minorities (Sproston & Mindell 2004), in addition to the use of own language, ethnic matching of the invitee and the person who maked the contact may promote participation (Font 2013)

Factor	Possible effects on participation rate
Incentives	The use of compensation or small "thank-you gifts" for participation (financial or other) may be considered. Prepaid incentives (paid with the survey itself) raise participation, while postpaid (paid after the survey) usually don't (Porter 2004). Long survey with incentives can make it achieve the same participation rate as a shorter survey without incentives (Groves 1999). The effect of incentives may depend on cultural norms.
Survey environment and background	Economic and social environments may affect by lowering or raising the participation rate; e.g. lower socio-economic groups tend to have lower participation rates (Harald 2007, Porter 2004).
Feedback from focus groups	Discussions in focus groups (small groups with representatives of potential participants) may produce important information for planning leaflets and invitations in a way that they raise interest to participate (Sinicrope 2009, Samanic 2003).
Internet survey vs. paper survey	Participation rate may be even higher in web survey compared to paper survey, but it depends on the population and the design of the web survey (Porter 2004, Link 2005). Typically web surveys can be used as an additional data collection method, since not all invitees have access to the internet (internet coverage varies between countries and population groups within countries). (See also Part A, Chapter 8. of the EHES Manual.)
Length of a questionnaire form	Long questionnaire forms (several pages) may have lower response rates than short forms (1-2 page), but only moderate effect (Porter 2004).
Deadline	Deadlines (giving respondents a deadline) haven't shown important effects on either increasing or decreasing the participation rate (Porter 2004).

The feasibility of personal contacts differs between countries and different population groups within the country, e.g. due to differences in the availability of telephone/mobile phone numbers and acceptability of home visits. Contacting potential participants by phone or mail may be challenging due to people's negative attitudes caused e.g. by numerous telemarketing calls and junk mail (Samanic 2003, Sinicrope 2009). Using media and different personal contact methods such as telephone calls, home visits, and reminders before appointment (phone call/ text message reminders) may help to raise the participation rate (Heistaro 2008, Tolonen 2014). Home visits are usually efficient in recruiting persons who are unable or unwilling to participate otherwise (Heistaro 2008, Lundqvist 2016). Participation should be facilitated through flexibility: re-scheduling of an appointment, prolonged opening hours, offering appointments also on weekends, possibility for selected persons to drop in without an appointment, and easy access to the examination site. Factors that may affect the participation rate are gathered in Table 13.1. It should be taken into account that the effect of some actions varies between cultures and sub-population groups and also within countries.

13.2.3.3 Partnership for enhancing participation

Partnership and collaboration with local organizations, professionals and communities help to raise awareness of the importance of the survey, and to arrange easy access to the examinations.

- The employers of the participants can be encouraged to allow their employees to participate in the survey during working hours.

- Cooperation with regional or local hospitals, non-governmental organizations, research centres and universities may increase the interest in participation.

- National and local health authorities and health professionals must be informed prior to the survey. They may help to motivate participation among their clients/patientes and they can help to report or to explain the measurements and laboratory results to the participants, making sure that there are personal benefits from participation.

- Local community leaders need to be notified to ensure the community's understanding and support.

- The public should be notified using mass media around the same time that the invitations are sent. (See Part A, Chapter 14.)

13.3 Non-participation

In order to assess the non-participation bias, it is important to collect information on non-participants to evaluate potential biases in estimates (Harald 2007, Jousilahti 2005, WHO MONICA Project 1997). This is important even when the participation rate is high. Some key information, such as age, sex and possibly some aspects of social status can in most countries be obtained already from the sampling frame or other registries through record linkage. In countries where it is possible to link the survey sample (also non-participants) to administrative registers such as hospitalisations, reimbursement of medications etc. register data can be used to obtain additional information from non-participants. In many countries, such record linkage is not possible and therefore, additional information needs to be obtained through non-participant questionnaire (see the non-participant questionnaire at the end of this Chapter). The questionnaire may be sent by mail or e-mail or mobile phone survey tool, or it can be filled in during a telephone interview or home visit. If the invited person is not available (by phone, e-mail or other means), proxy information may be used for completing the short non-participant questionnaire. The mode of the data collection and the use of a proxy should be recorded (see Part A, Chapter 8).

13.4. Adjustment methods for non-participation bias

Non-participation bias can never be corrected with statistical methods, therefore it is essential to try to obtain as high participation rates as possible. With statistical methods we can estimate potential effects of non-participation on our results. There are several statistical methods which can be used for estimation under missing data due to non-participation. Methods such as expectation-maximation (Dempster 1977, Scheike 2004), Bayesian data augmentation (Kulathinal 2006), con-

ditional likelihood (Saarela 2009, Saarela 2012), parametric or non-parametric multiple imputation (Karvanen 2010, Rubin 1987, Zhou 2001), propensity scores (Little 1988, Rosenbaum 1983), and weighting methods (Breslow 2007, Gray 2009, Horvitz 1952, Samuelsen 2007, Särndal 1992) can be used. Access to auxiliary information on non-participants through sampling frames, non-participation questionnaires and possibly also through record linkage to the administrative registers are important for use of these methods.

13.5 Data to be recorded on recruitment process

It is necessary to keep a record of the participation status of each person invited to participate in the HES. The number and type of contact attempts should be recorded. If the person was contacted, it should be recorded if the person participated, refused or dropped out after having agreed to participate. Information on completed and not completed examinations and questionnaires should be recorded. If the person refused, the reason should be recorded, if this information can be obtained. Reasons for not being examined are listed below. Some of these reasons should not only be recorded but it should be attempted to convince the person that his/her participation is highly valued.

Reasons for not being examined:

- Refused: no reason given
- Refused: lack of time
- Refused: personal principle
- Refused: health problem (e.g. disability restricting access to the examination site or is hospitalised)
- Refused: feeling healthy (therefore thinks that there is no reason to participate)
- Refused: survey topic (is not interested in health issues or considers this too personal)
- Contacted: not able to schedule an appointment (e.g. participant could only attend during evening hours or week ends)
- Contacted: not showing up (does not come to the scheduled visit, and the visit cannot be re-scheduled)
- Not contacted: not reached (no address/phone number available, outdated information)
- Not eligible: moved abroad
- Not eligible: moved out of the primary sampling unit (PSU)
- Not eligible: age out of survey range
- Not eligible: died
- Temporarily unavailable: e.g. holiday, working or studing outside the survey area for a long time

- Language problems
- Impossible to examine for other reason (this reason should be specified, if feasible)

References

- Aromaa A, Koponen P, Tafforeau J, Vermeire C, (eds.). Status and future of health surveys in the European Union. Final report of Phase 2 of the project Health Surveys in the EU: HIS and HIS/HES Evaluation and Models. Helsinki: National Public Health Institute; 2003.

- Barchielli A, Balzi D. Nine-year follow-up of a survey on smoking habits in Florence (Italy): higher mortality among non-respondents. Int J Epidemiol 2002 Oct; 31(5): 1038-42

- Breslow NE, Wellner JA. Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. Scandinavian Journal of Statistics. 2007;34(1):86-102.

- Cohen C, Duffy JC. Are nonrespondents to health survey less healthy than respondents? Journal of Official statistics 2002;18(1):13-23

- Corbin-Smith G, Canty Williams I, Blumenthal C et al. Relationships and communication in minority participation in research: multidimentional and multidirectional. J Nat Med Assoc 2009; 99(5): 489-498

- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, B. 1977;39(1):1-38.

- 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 Jan 1; 147(1): 74-82

- Edwards PJ, Roberts I, Clarke MJ et al. Methods to increase response to postal and electronic questionnaires. Cochraine Database of Systematic Reviews 2009, Issue 3, Art. No. MR000008. doi: 10.1002/14651858. MR000008.pub4

- European Health Interview & Health Examination Surveys Database. 2010 [cited 7 April 2010];Available from: https://hishes.wiv-isp.be

- Font J, Mendez M. Surveying Ethnic Minorities and Immigrant Populations. Amsterdam University Press, Amsterdam 2013.

- Gray RJ. Weighted analyses for cohort sampling design. Lifetime Data Analysis. 2009;15(1): 24-40.

- Groves RM, Singer Eleanor, Corning AD, Bowers A. A laboratory approach to measuring the effects on survey participation of interview length, incentives, differential incentives, and refusal conversion. Journal of Official Statistics 1999;15(2):251-268.

- Hara M, Sasaki S, Sobue T, Yamamoto S, Tsugane S. Comparison of cause-specific mortality between respondents and nonrespondents in a population based prospective study: ten-year follow-up of JPHC study cohort I. Japan Public Health Centre. J Clin Epidemiol 2002 Feb; 55(2):150-6

- Harald K, Salomaa V, Jousilahti P, Koskinen S, Vartiainen E. Non-participation and mortality in different socioeconomic groups: the FINRISK population surveys in 1972-92. Journal of epidemiology and community health. 2007 May;61(5):449-54.

- Heistaro S. Methodology report. Health 2000 Survey. KTL-National Public Health Institute, Finland. Helsinki 2008.

- Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 1952;47:663-85.

- Jackson R, Chambless LE, Yang K et al. Differences between respondents and nonresponsents in a multicentre community-based study vary by gender ethnicity. The Atherosclerosis Risk in Communities (ARIC) Study Investigators. J Clin Epidemiol 1996 Dec; 49(12): 1441-46

- Jousilahti P, Salomaa V, Kuulasmaa K, Niemela M, Vartiainen E. Total and cause specific mortality among participants and non-participants of population based health surveys: a comprehensive follow up of 54 372 Finnish men and women. Journal of epidemiology and community health. 2005 Apr;59(4):310-5.

- Kaila-Kangas L, (ed.) Musculosceletal Disorders and Diseases in Finland. Results of the Health 2000 Survey. Publications of the National Public Health Institute, B25/2007. Helsinki, 2007

- Karvanen J, Saarela O, Kuulasmaa K. Nonparametric multiple imputation of left censored event times in analysis of follow-up data. Journal of Data Science. 2010;8:151-72.

- Koponen P, Aromaa A. Survey design and methodology in national health interview and health examination surveys review of literature, European survey experiences and recommendations. In: Aromaa A, Koponen P, Tafforeau J, Vermeire C, Primatesta P, Marmot M, et al., eds. Health surveys: evaluation and recommendations Subproject reports of Phase 2 of the project Health Surveys in the EU: HIS and HIS/HES Evaluation and Models. Helsinki: National Public Health Institute 2003.

- Kulathinal S, Arjas E. Bayesuan inference from case-cohort data with multiple end-points. Scandinavian Journal of Statistics. 2006;33:25-36.

- Little RJA. Missing-data adjustments in large surveys. Journal of Business and Economic Statistics. 1988;6:287-96.

- Link MW, Mokdad AH. Alternative modes for health surveillance surveys: an experiment with web, mail, and telephone. Epidemiology. 2005 Sep;16(5):701-4.

- Lundqvist A, Mäki-Opas T. Health 2011 Survey – Methods. National Institute for Health and Welfare, Report 8/2016. Available at http://urn.fi/URN:ISBN:978-952-302-669-8.

- Macera CA, Jackson KL, Davis DR, Kronenfeld JJ, Blair SN. Patterns of nonresponse to a mail survey. J Clin Epidemiol 1990; 43(12): 1427-30

- Mindell J, Giampaoli S, Goesswald A et al. Sample selection, recruitment and participation rates in health examination surveys in Europe – experience from seven national surveys. BMC Med Res Methodol 2015; 15:78. doi:10.1186/s12874-015-0071-4

- Phillips M, Woodward C, Collins D, O'Connor W. Encouraging and maintaining participation in the families and children survey: Understanding why people take part. 2002. Working paper number 6. National Centre fo Social Research. Available from: http://research.dwp.gov.uk/asd/asd5/WP6.pd

- Porter AR. Raising Response Rates: What works, New Directions for Institutional Research 2004;121:5-21

- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41-55.

- Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wile & Sons, Inc.; 1987.

- Saarela O, Kulathinal S, Karvanen J. Joint analysis of prevalence and incidence data using conditional likelihood. Biostatistics. 2009;10:575-87.

- Saarela O, Kulathinal S, Karvanen J. Secondary analysis under cohort sampling designs using conditional likelihood. Journal of Probability and Statistics. 2012;47.

- Samanic CM, Colt J, Campbell W, WH Chow. Attitudes towards participation in a population based epidemiologic study: Results from focus groups. Annals of Epidemiology 2003;13(8):559.

- Samuelsen SO, Anestad H, Skrondal A. Stratified case-cohort analysis of general cohort sampling designs. Scandinavian Journal of Statistics. 2007;34(1):103-19.

- Scheike TM, Martinussen T. Maximum likelihood estimation for Cox's regression model under case-cohort sampling. Scandinavian Journal of Statistics. 2004;31:283-93.

- Sharar E, Folsom AR, Jackson R. The effect of nonresponse on prevalence estimates for a referent population: insights from a population-based cohort study. Atheresclerosis Risk in Communities (ARIC) Study Investigators. Ann Epidemiol 1996 Nov; 6(6): 498,506

- Sinicrope PS, Patten CA, Bonnema SM, Almquist JR, Smith CM, Beebe TJ, Jacobsen SJ, Vachon CM. Healthy women's motivators and barriers to participation in a breast cancer cohort study: a qualitative study. AEP 2009;19(7):484-493.

- Sproston K and Mindell J (eds). Health Survey for England 2004. Volume 2: The health of minority ethnic groups. Methodology and documentation. The Information Centre, 2006a.

- Spry VM, Hovell MF, Sallis JG, Hofstetter CR, Elder JP, Molgraad CA. Recruiting respondents to mailed surveys: Controlled trials of incentives and prompts. American Journal of epidemiology 1989;130(1):166-172.

- Särndal C-E, Swensson B, Wretman J. Model assissted survey sampling. New York: Springer; 1992.

- Tolonen H. Towards the high quality of population health surveys. Standardization and quality Control. National Public Health Institute A27/2005, Helsinki 2005 [cited 1 January 2010]; Available at http://urn.fi/URN:ISBN:951-740-546-4. Accessed on 22 January 2013

- Tolonen H, Dobson A, Kulathinal S. 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(11):887-98.

- Tolonen H, Kuulasmaa K, Laatikainen T, Wolf H. Recommendations for indicators, international collaboration, protocol and manual of operations for chronic disease risk factor surveys. EHRM Project. 2002 [cited 1 January 2010]; Available a http://www.thl.fi/publications/ehrm/product2/title.htm. Accessed on 22 January 2013

- Tolonen H, Aistrich A, Borodulin K. Increasing health examination survey participation by SMS reminders and flexible examination times. Scand J Public Health 2014; 42(7): 712-7

- Tolonen H, Ahonen S, Jentoft S et al. Differences in participation rates and lessons learned about recruitment of participants – the European Health Examination Survey Pilot Project. Scan J Public Health 2015; 43(2): 212-9. doi:10.1177/1403494814565692

- Tupasela A, Sihvo S, Snell K, Jallinoja PA, Aro AR and Hemminki E. Attitudes towards biomedical use of tissue sample collections, consent, and biobanks among Finns. Scand J Public Health 2010; 38; 46 originally published online Nov 11, 2009;DOI: 10.1177/1403494809353824.

- van den Akker M, Buntinx F, Metsemakers FK, Knottnerus JA, Morbidity in responders and non-responders in a register-based population survey. Fam Pract 1998 Jun; 15(3): 261-3

- WHO MONICA Project. MONICA Manual, Part III: Population Survey, Section 1: Population survey data component. 1997 [cited 01 January 2010]; Available at http://www.thl.fi/publications/monica/manual/part3/iii-1.htm. Accessed on 22 January 2013

- Wolf HK, Kuulasmaa K, Tolonen H, Sans S, Molarius A, Eastwood BJ. Effect of sampling frames on response rates in the WHO MONICA risk factor surveys. European journal of epidemiology. 2005;20(4):293-9.

- 52. Zhou X-H, Eckert GJ, Tierney MW. Multiple imputation in public health research. Statistics in Medicine. 2001;20:1541-9.

Examples of the information leaflet, invitation letter and non-participant questionnaire

EHES

Health Examination Survey

Thank you for your help with this survey. Your co-operation is very much appreciated.

For further information, please contact:
Name of the contact person
Address
Toll-free phone number
E-mail address
Survey web page

EHES
Measuring the Health of Europeans

What is this survey about?

The intention of this health examination survey is to receive up to date health information of the adult population of x (*country*). The information gathered will be used for planning health care as well as assessment of the prevalence of diseases, their causes, and care. The survey is being carried out by x and y (*name of the organization/partners*).

Why am I selected?

We have invited 4000 people to take part in the survey. You have been randomly selected from the national population register.

Why is it important to participate?

This survey is important for improving public health. As you have been selected to the sample, your participation is very important. As the selection has been made by random, it is not possible to replace a selected person by anyone else.

Do I benefit from the survey?

Yes, you will receive important information about your health. During the examinations you have a change to get feedback on results and talk to the health professional. After the examination you will get a report of your results in mail. The health examination is free of charge.

All participants will receive xx (*if incentives are used*).

What measurements are included?

The measurements include height, weight, waist circumference, blood pressure and x (*list additional measurements*). Also a blood sample will be taken to measure total and HDL-cholesterol and glucose.

The measurements are safe and are made by specifically trained and qualified personnel.

Is the survey confidential?

All survey data is confidential and protected by legislation (Data Protection Act). This means that survey results will not be presented to reveal your identity at any point.

Is the survey compulsory?

Participation is completely voluntary. The success of the survey relies on the co-operation and goodwill of those asked to take part. The more people take part, the more useful the results are. You may withdraw from the survey at any point.

Whom can I contact to ask further questions?

We will help you with any questions or concerns you may have. Please call us at xx-xx-xxx (*toll-free phone number*). The survey website at http://www.hes.xx also has more information.

28 April 2011   *Study ID*

Mr./Ms. First name Last name
Street address
City

Dear Mr. /MS. Last name,

We are inviting you to participate to the Health Examination Survey of country x (*substitute with the survey name*). This survey studies the health of population in country x (*replace with your country*). You have been selected from national population register to represent 25-64 years old people of the country (*replace with your country*).

In the survey, an interview will be conducted and your height, weight, waist circumference, blood pressure will be measured and blood sample collected.

Representativeness and usefulness of the results of the survey depend on people we contact to get involved. It takes 30-45 minutes to go through the interview and measurements. You cannot be replaced by anyone else. Your participation is voluntary.

All information collected during the survey, will be handled confidentially. You can find answers to the questions regarding the survey from attached leaflet. You can also call on Monday-Friday at 9:00-16:00 to TOLL-FREE-PHONE-NUMBER if you have any questions.

Our survey team will contact you within next few days to arrange the appointment time for you.

The HES survey team thanks you for your collaboration.

Sincerely,

_____   _____
Mark Model, Dr.   Susie Super, PhD
Project Leader   Head of Department

28 April 2011   *Study ID*

Mr./Ms. First name Last name
Street address
City

Dear Mr. /MS. Last name,

We are inviting you to participate to the Health Examination Survey of country x (*substitute with the survey name*). This survey studies the health of population in country x (*replace with your country*). You have been selected from national population register to represent 25-64 years old people of the country (*replace with your country*).

In the survey, an interview will be conducted and your height, weight, waist circumference, blood pressure will be measured and blood sample collected.

Representativeness and usefulness of the results of the survey depend on people we contact to get involved. It takes 30-45 minutes to go through the interview and measurements. You cannot be replaced by anyone else. Your participation is voluntary.

All information collected during the survey, will be handled confidentially. You can find answers to the questions regarding the survey from attached leaflet.

We have booked you an appointment for the examination clinic (*provide address of the clinic*) on

6 May 2011 at 8:30.

If this time is not suitable for you, please call on Monday-Friday at 9:00-16:00 to TOLL-FREE-PHONE-NUMBER to schedule new appointment.

Please, read the instructions to the participants leaflet attached to this invitation before coming to the examination clinic.

The HES survey team thanks you for your collaboration.

Sincerely,

_____   _____
Mark Model, Dr.   Susie Super, PhD
Project Leader   Head of Department

Non-participant questionnaire

Version: 22 March 2011

Identification

Participants identification code:

Background information

Sex

- ☐ Man
- ☐ Woman

Date of birth (dd.mm.yyyy)

Age (in full years)

Educational level

What is the highest education leaving certificate, diploma or education degree you have obtained? (Please, include any vocational training)

- ☐ No formal education of below ISCED 1
- ☐ Primary education (ISCED 1)
- ☐ Lower secondary education (ISCED 2)
- ☐ Upper secondary education (ISCED 3)
- ☐ Post-secondary but not-tertiary education (ISCED 4)
- ☐ First stage or tertiary education (ISCED 5)
- ☐ Second stage of tertiary education (ISCED 6)

Health status

How is your health in general?

- ☐ Very good
- ☐ Good
- ☐ Fair
- ☐ Bad
- ☐ Very bad

Diagnosed diseases

Do you have or have you ever had any of the following diseases or conditions, diagnosed by a medical doctor?

Myocardial infarction	☐ Yes	☐ No
Coronary heart disease (angina pectoris)	☐ Yes	☐ No
High blood pressure (hypertension)	☐ Yes	☐ No
Evelated blood cholesterol	☐ Yes	☐ No
Stroke	☐ Yes	☐ No
Diabetes	☐ Yes	☐ No

Height and weight

How tall are you without shoes? (cm)

How much do you weight without clothes and shoes? (kg)

Smoking status

Do you smoke at all nowadays?

- ☐ Yes, daily
- ☐ Yes, occasionally
- ☐ Not at all

Reason for non-participation

Why did you not participate to the survey?

- ☐ Not interested
- ☐ No time
- ☐ Not able to get suitable appointment time
- ☐ I'm healthy, no need to participate
- ☐ I'm too ill to participate
- ☐ Don't participate to any surveys

14. Dissemination and publicity

PART A

Hanna Tolonen[1], Päivikki Koponen[1]

[1] National Institute for Health and Welfare (THL), Helsinki, Finland

14.1 Purpose of dissemination

Dissemination and reporting are needed to increase knowledge, instruct, facilitate informed decision making, and persuade. (Nelson 2009) In a national HES this would translate to actions related to:

- Seeking the support of the decision makers to carry out the HES. This will require justification of the aims of the HES for funders, and demonstrating the benefits of the HES for different stakeholders.

- Seeking the support of the other stakeholders such as non-governmental organizations (e.g. hypertension and diabetes associations) which may provide technical and scientific support for the development of survey contents and protocols.

- Seeking the support and collaboration of local authorities, health and social care organizations etc. to obtain their collaboration and promotion of the survey.

- Motivating invitees to participate and informing them about the survey and requirement for the measurements.

- Informing participants about their personal measurement results and, if feasible, also about the overall results of the survey.

- Dissemination of the key findings for the general public, policy makers, health care authorities, non-governmental organizations, scientific community, etc. to promote utilization of the survey findings and as a basis for future surveys and specific research projects

In health examination surveys, the availability of the basic information about the survey is critical. Easy access to high quality information will assert confidence in the survey and focus the attention of stakeholders.

14.2 Dissemination plan

A dissemination plan should be prepared at an early stage of the survey process. It will provide structured framework for the dissemination in the different phases of the survey process. The dissemination plan will define what, when, to whom, how and by whom different matters regarding the HES will be disseminated/reported. The structure of the dissemination plan could be as follows:

- the purpose of the dissemination in the national HES (see 14.1);
- target groups for dissemination;
- key messages to be disseminated;
- means of dissemination (including tools, timing, and responsibilities);
- resources; and
- evaluation of the dissemination process and outcome.

14.2.1 Target groups

It is important to identify relevant target groups for the dissemination. In a national HES, target groups may vary in different phases of the survey. There may also be variation between surveys and countries depending on the contents of the survey, national legislation etc. Table 14.1 provides a tentative list of potential target groups in different phases of a HES.

Table 14.1 Potential target groups for dissemination

Survey phase	Target group
Planning	Policy makers (ministries)
	Health and social care authorities (regional health authorities, heads of local health care centres, social services, etc.)
	Non-governmental organizations
	Research community
	Possible other funders and stakeholders
Fieldwork implementation	General public
	Survey invitees
	Health and social care authorities, and local health and social care personnel
	Other local authorities
	Media

Survey phase	Target group
Survey results	Survey participants
	General public
	Policy makers
	Non-governmental organizations
	Health and social care authorities and personnel
	Research community
	Funders
	Media

During **the planning phase** it is important to provide information for those whose support for a HES is needed. Support can be on political level (national/regional), providing knowledge to the development of survey contents and methods, financial support, or as promotion of the survey. At the national level there may also be other authorities whose permission/authorization is required.

During **the fieldwork implementation phase** the most important target groups are the survey invitees and general public. Their collaboration is the key for the success of the survey. Other target groups such as different authorities are often needed also during the fieldwork to support the fieldwork locally.

After the survey is finished and **survey results** are ready for reporting, it is important to provide information for all stakeholders. The level and format of information may vary between target groups (see Section 14.2.3).

Each target group has its own function in the survey process:

- Policy makers can be from national, regional or local level. They are end users of the survey results and may utilize results such as prevalences and changes in risk factors and health behaviours in the population for development of evidence-based policy. Their support is usually needed when making the decision to conduct a national HES. The support can be political and financial.

- Health and social care authorities and personnel at the local level will provide support for the survey. Health and social care authorities are users of the survey results and they can also provide material/financial support for the survey. It is important to keep health and social care personnel, especially general practitioners/physicians informed about the survey since invitees may contact them to learn more about the survey before their final decision on participating. General practitioners/physicians can also provide follow-up and additional examinations and care in the case of abnormal results in the survey.

- Local authorities can support the survey and ensure secure environment for the fieldwork. For example, if examinations are conducted by home visits it may be important to inform the policy force about the survey in case people call police to

report about strangers on their doorsteps. Local authorities may also help in finding and getting access to the premises to be used in the health examinations, and support in the general positive attitudes to enhance participation.

- Non-governmental organizations (NGOs) are often users of the survey results but they can also support survey planning. During the survey planning, NGOs may have special expertise on some of the topics/measurements planned for the survey and can provide support on preparation of the survey protocol and training for these topics/measurements. They may also provide financial support and help to raise awareness of the general public on the survey through their networks.

- Research community has an important role in planning and reporting on the results. During the survey planning, knowledge of research community is needed for the preparation of the survey contents and protocols, and organizing training. Research community is an important user of the collected survey data and can sometimes also provide material and/or financial support.

- General public and the survey invitees are key partners in the HES. Their collaboration is vital for obtaining a high participation rate and therefore reliable and representative survey results. Even though only persons in a random sample are invited, it may be important to inform the general public to promote general awareness about the survey. This may lead to positive attitudes and interest in the survey, which can help when the person gets the invitation. The invitees' family members and employers may also be more willing to support his/her choice to participate if they are aware of the survey.

- Media has a key role delivering information about the HES to the general public and other stakeholders.

14.2.2 Key messages

The key message for dissemination depends on the target group and phase of the survey. The message should be clear, simple and easy to understand, tailored to the recipient(s), correct and realistic. The messages should be simple and consistent.

During the planning phase of the survey, key messages relate to the importance of HESs for health monitoring, public health planning and evaluation, and the research potentials of the survey data. Examples of messages:

- Health surveys are vital for understanding the health situation and the behaviours of the population, and they provide an evidence-base for health policies.

- Identifying health disparities between population groups is a prerequisite for narrowing down health inequalities.

- To support healthy aging and to prevent early retirement we need to know the current state of health of the adult working aged population.

During the fieldwork phase, key messages are targeted more at survey invitees to motivate them to participate and at the general public to support participation. Examples of messages:

- The national HES is conducted by a reliable public health authority, the methods are secure and science based, and the results do not serve any other interests but the public benefit.

- Participants will have a free-of-charge opportunity to receive up-to-date information on their own health.

- Information about people's health is vital to building up an efficient health care system geared to our health needs and that of our families. Each individual's contribution is important in making the study representative.

- All information collected is handled following the data protection legislation. You cannot be identified from any of the reported results.

- After the survey has been finished and results are ready for dissemination, messages have to be tailored for different target groups. Survey participants are often provided at least with their own results and this report may also include some population level summary results. Policy makers and health care administrators, and many other stakeholders, need concise messages with relevant interpretation. Researchers, on the other hand, need detailed information about the survey and the collected data and the procedures for getting access to the data for further research.

14.2.3 Means

14.2.3.1 Brand building

It is important to pay attention to how the survey is presented in all dissemination. This is a kind of brand building of the national HES. The brand of the survey is used to identify the survey from others.

Key points for building the brand for a national HES are:

- Identify what qualities, values and expertise the HES is associated with.

- Think like the 'customer' of the HES. What kind of impression do you what them to get?

- Select a slogan and/or few main phrases to be associated with the HES.

- Select a distinctive visual layout/picture.

- Get your entire survey team involved in using the brand.

- Promote your brand.

The brand building aims to provide a uniform and professional message and image about the survey.

The tools for the brand building of a national HES are the name of the survey and its acronym, logo, slogan(s), website and all promotion materials. The same selected survey name, logo, slogan, colors and font should be used in all survey and promotion materials such as the template for presentations (e.g. PowerPoint), templates for other published materials (e.g. letters, posters, questionnaires, press releases etc.), signs in the examination sites, pens, notepads, cups, etc. and also in the appearance of the fieldwork staff (name tags, possibly also clothing).

Everyone working on the survey should be aware about the brand and related material and use them actively in their everyday work.

14.2.3.2 Tools

There are different methods for dissemination of the information about a HES (Table 14.2).

Table 14.2 Methods for dissemination

Method	Purpose
Face-to-face contact	Engage
	Increase awareness
	Inform
Workshop/seminar	Increase awareness
	Inform
	Promote
Leaflets and brochures	Increase awareness
	Promote
Newsletters	Increase awareness
	Promote
Personal letters	Engage
	Increase awareness
	Inform
	Promte
Press releases and press conferences	Increase awareness
	Promote
Website	Increase awareness
	Inform
	Promote
Social media	Increase awareness
	Promote

Method	Purpose
Reports	Increase awareness
	Inform
	Promote
Scientific journal articles	Increase awareness
	Promote
Conference posters/presentations	Increase awareness
	Prmote
TV and radio advertisement	Increase awareness
	Promote
Information desks on public places	Engage
	Increase awareness
Promotion materials	Increase awareness

- Engage: obtain personal engagement to the survey.
- Awareness: increase awareness about the survey.
- Inform: provide detailed information for specific target groups.
- Promote: promoting results of the survey.

See for example Hall (1998) and Froyd (2001) for more information about dissemination strategies and some examples on how different dissemination strategies can be used in survey setting.

The right method(s) need to be selected for each message and target group. In this selection, stakeholder analysis can also be helpful. Stakeholder analysis can help to identify stakeholders who need or may benefit from specific contacts. In stakeholder analysis you identify those who are likely to affect or be affected by HES, and sort them according to their impact on action and the impact the action will have on them. This can be done in many ways (Mitchell et al 1997, Fletcher et al 2003, Cameron et al 2010, Savage et al 1991 and Turner et al 2002). One of the often used methods is the influence(power)-interest grid where stakeholders are classified to a matrix (Figure 14.1).
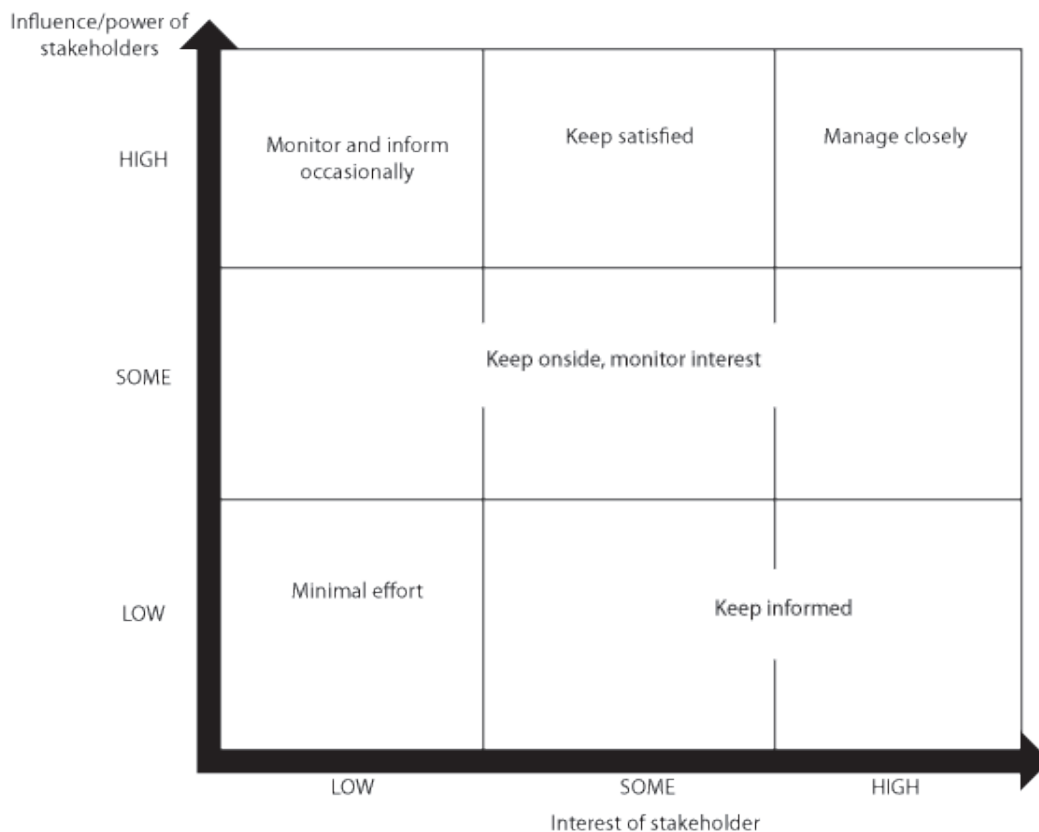
Figure 14.1 Stakeholder power and interest matrix

14.2.3.3 Timing

Right timing of the dissemination is important for its success. Messages should be timed according to the event they are related to. For example, promotion of the start of the survey should be within few days from the starting date, not months before when people will already forget it by the time they get an invitation.

It may happen occasionally that the survey organizers cannot control the optimal timing. There may be positive or negative unintentional publicity which has to be responded to when it happens. For example, an individual participant may report about an unpleasant (or pleasant) survey experience in social media and this gets wide attention. Perhaps the printed media picks this up and contacts the survey organizers for a response to the accusations or other feedback.

There should be a plan on how to react and by whom in this type of situations. Open communication is the key. The response should be provided without delay, based on facts and avoiding unnecessary explanations.

14.2.3.4 Responsibilities

The dissemination plan should designate for each task/action a person, who makes sure that the task/action gets done (what, to whom,

how and when). This person is also responsible for reporting if, for any reason, the task/action is delayed or cannot be completed as planned.

14.2.4 Resources

All dissemination activities require resources, either personnel and/or material, which need to be included in the survey budget. The extent of the dissemination activities will depend on the available funds but usually at least the following should be budgeted: personnel costs, layout and printing of the materials, translation of the materials, web domains, promotion materials, equipment (stands, projectors, etc.), travel costs, advertisements in different media, software/licenses, and room rentals and catering for workshops and seminars. Sometimes it is possible to obtain 'free' (unpaid) publicity for example in local radio stations, social media, newspapers and on TV. Such opportunities should be looked for whenever possible.

A check list of issues to be included in the dissemination budget is provided in Part A, Chapter 16.

14.2.5 Evaluation

In each survey, someone should be responsible for the follow-up of the implementation of the dissemination plan, monitoring media coverage and evaluation of the progress and outcomes. For each dissemination activity, measurable criteria should be set already at the beginning of the survey process when the dissemination plan is prepared. The main focus should be on outcomes. Prefer quality of your dissemination over quantity.

Dissemination plan can be updated during the survey process if needed.

14.3 Example dissemination plan

In a simple format, the dissemination plan can be a table like this.

Action/ Dissemina- tion	Why?	To whom? (Targtet group)	What? (Message to be disseminated)	How? (Tools)	When? (Timing)	Who? (Person respon- sible for the task)	Resources	Evaluation

References

- Cameron BG, Seher t, Crawley EF. Goals for space exploration based on stakeholder network value considerations. Acta Astronomica 2010, doi: 10.1016/j.actaastro.2010.11.003

- Fletcher A et al. Mapping stakeholder perceptions for a third sector organization. Journal of Intellectual Capital 2003;4(4):505-527

- Froyd J. Developing a dissemination plan. 31st ASEE/IEE Frontieres in Education Conference, 10-13 October 2001

- Gehrke D, Turban E. Determinants of successful website design: relative importance and recommendations for effectiveness. Proceedings of the 32nd Hawaii International Conference on System Sciences, 1999, pp.1-8

- Hall MH. A Dissemination strategy for the national survey of giving, volunteering and participating. Canadian Centre for Philanthropy, 1998. Available at: http://sectorsource.ca/sites/default/files/resources/files/nsgvp_dissemination_strategy.pdf

- Mitchell RK, Agle BR, Wood DJ. Towards a theory of stakeholder identification and salience: defining the principle of who and what really counts. Academy of Management Review 1997;22(4):853-888

- Nelson DE, Hesse BW, Croyle RT. Making data talk. Communicating public health data to the public, policy makers, and the press. Oxford University Press, 2009

- Savage GT, Nix TW, Whitehead, Brair. Strategies for assessing and managing organizational stakeholders. Academy of Management Executive 1991; 5(2): 61-75

- Turner JR, Kristoffer V et al. (Eds.) The project manager as change agent. London. McGraw-Hill Publishing Co. 2002

15. Training programme

PART A

Päivikki Koponen[1], Hanna Tolonen[1], Kari Kuulasmaa[1]

[1] National Institute for Health and Welfare (THL), Helsinki, Finland

This chapter outlines the recommended training programme and presents key issues which should be considered when planning and preparing the national training programmes for all staff members who take part in the data collection. It is essential to outline the national training at early stages of the planning process, as this will affect both budgeting (training costs) and timing of the data collection. Training is a key element of standardization and quality assurance (see Part A, Chapter 11). Training material related to the core measurements is available at the EHES web site (http://www.ehes.info/rc/training_seminar/training_seminars.htm) to be used and adapted freely for national training.

15.1 EHES training

The EHES training programme includes two dimensions:

1. Europe wide training seminars for the persons responsible for the planning and organizing the surveys and for those resposible for training of the national survey personnel, and

2. Outline for the training of the national survey personnel actually conducting the survey. This training is conducted nationally.

15.1.1 EHES training seminars

European training seminars should be organized periodically, for all countries planning and preparing their HES. The possibilities to organize such seminars will depend on the resources for the EHES RC in the future.

Three training seminars are recommended with the aims to ensure standardization and to share experiences between countries.

1) Training seminar covering issues **relating to the planning and preparing** for the European Health Examination Survey (EHES) at the national level.

> The target group for this seminar is those who plan and prepare national surveys in Europe. The objective of the seminar is to promote planning national HESs according to the EHES standards. Other objectives of the seminar are to raise awareness on EHES in all European countries, to receive feedback on the EHES standards and the EHES Manual, and to discuss possible national adaptations, and if feasible, to discuss and share experiences on potential additional measurements. The seminar will support preparing the national manuals and finalizing national study plans especially for counties without previous experiences in organizing HES but also experienced survey organizers may benefit from sharing experiences.

2) Training seminar covering issues relating to **the fieldwork** of the national health examination surveys.

> The target group for this seminar is those who will train the national fieldwork team members in each country. The objective of the seminar is to promote the use of the standard EHES training materials and to otherwise ensure that the training for the fieldworkers will be organized following the EHES standards. The focus is on the core measurements but also additional measurements can be included, when feasible. The seminar will support finalizing the national manuals and training programmes especially for counties without previous experiences in organizing HES but also experienced survey organizers may benefit from sharing experiences and promoting standardization and comparability of results between countries.

3) Training seminar focusing on **data management, validation and analysis, and reporting and dissemination of results**.

> The target group of this seminar is ICT-personnel, statisticians, researchers and survey organizers responsible for the data management, validation and analysis and reporting and dissemination of the results. The aim is to promote comparison of the national results, to develop European level reports and to support both national and European dissemination of the results.

15.1.2 EHES training materials

Some training materials for EHES are available at the EHES website http://www.ehes.info. The national survey organizers and national trainers are encouraged to translate, use, develop further and adapt these materials for their national purposes. However, they should keep in mind that the key contents and methods for the national training should be standardized to assure the international comparability.

15.2 National training programme

All members of the national survey team, both those working at the central office and all fieldwork staff members should participate in the national training. It is essential for the quality of the survey that everyone, including secretaries and assistants working at the central survey office, those who contact the selected persons, send the invitations and schedule the visits, data managers, statisticians and all field work staff members know and understand the aims of the survey and the whole data collection process.

The key contents of the national training should be similar in all countries, but some parts will depend on how the fieldwork is organized and if other measurements are carried out in addition to the EHES core measurements.

15.2.1 Outline for the national training seminars

The training can be divided in the following modules:

1. for all staff members;
2. for personnel at the central office;
3. for fieldwork personnel. (Figure 15.1).
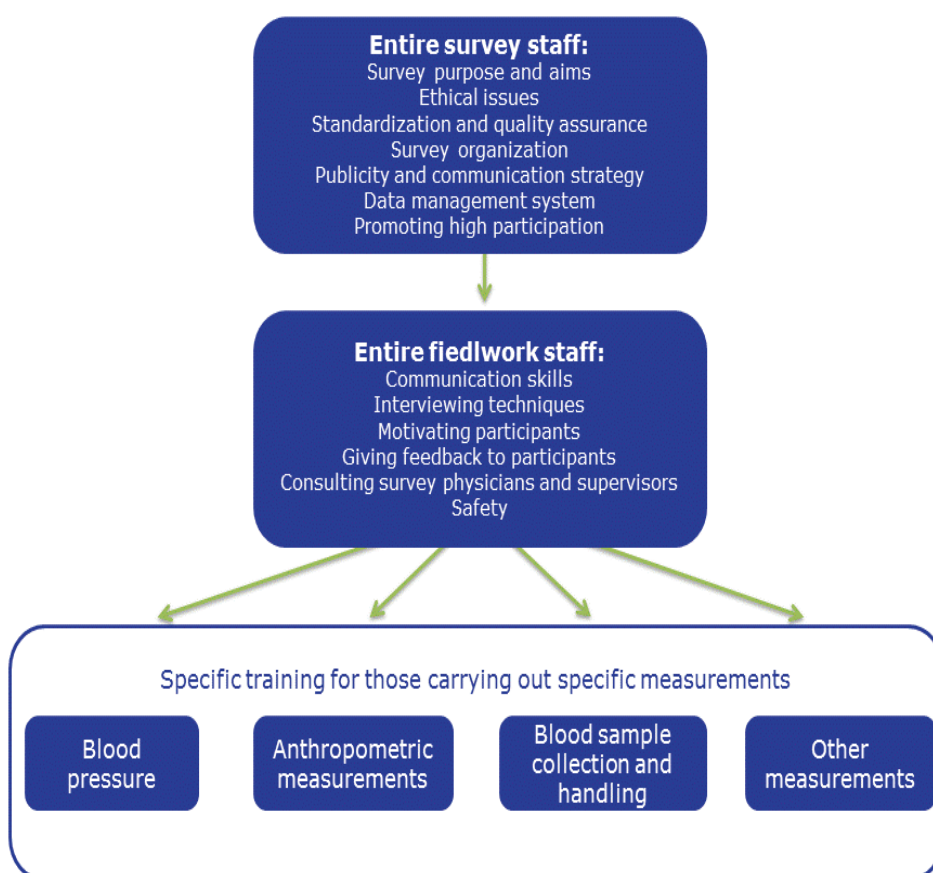
Figure 15.1 Training process for survey staff members

If the staff members have experience from previous surveys some parts of the general training may be only short refresher lectures. Practical measurement sessions are needed also for the experienced staff members to ensure that the standards are followed correctly.

The training should include at least the following topics for **all staff members**:

- Purpose and aims of the survey
  - It is important that all staff members understand the importance of the survey and are able to describe the aims and purpose of the survey to the participants in a standard way;
- Ethical issues and confidentiality
  - What is data confidentiality and how it is assured by all staff members, why an informed consent is needed, what is meant by the informed consent, and how the informed consent should be obtained;
- Random samples and the importance of high participation rates
  - How people are selected, and why all selected persons are equally important regardless of their health status or other characteristics, how participation can be encouraged and motivated;
  - How people are invited;
- The importance of standardization and quality assurance
  - Understanding the aims and the role of the survey manual, audit visits and quality assurance, the importance of consulting supervisors when needed;
- Survey organization
  - roles and responsibilities of each staff member at the central office and in the fieldwork teams;
- Communication skills
  - including similarities and differences in professional conduct during survey data collection and clinical practice in normal health care settings;
- Dissemination and reporting
  - How and when the survey results will be reported and published, publicity rules and working with local media during fieldwork;
  - How the participants will be informed about their personal examination and laboratory results;
- Data management
  - The data management system and separate sesssionsfor specific ICT skills needed at the central office and in the fieldwork for data entry, handling and reporting.

Training for the **personnel at the central office:**

- Specific tasks at the central office such as e.g. recruitment of selected persons, mailing invitations and personal results, scheduling appointments, recording contacts with invitees

Training for the **fieldwork staff**:

- Local collaboration during fieldwork
    - Working with the local health care professionals e.g. to build and maintain good collaboration, so that they encourage their patients to participate in the survey, and referring participants with abnormal measurement results to their GPs or other local health care professionals;
- Interviewing and checking self-administered questionnaires;
- As in most cases all fieldwork staff members have at least some interview questions to be asked before or after the clinical measurements, all of them will need training in general interviewing skills, including (adapted from Czaja & Blair 2004):
    - rules for accepting proxy responses;
    - reading questions verbatim;
    - using non directive probes (when allowed);
    - asking all questions;
    - recording answers correctly, especially in case of open ended questions;
    - specific procedures for each interview module or instrument.
- Specific measurements:
    - rationale why they are measured,
    - measurement techniques,
    - including practical training and certification if needed,
    - giving feedback to participants concerning measurement results;
- Consulting survey physicians and local health care professionals when needed;
- Safety of the fieldwork team members (e.g. actions needed in case of needle stick injuries, violently acting and aggressive participants);
- ICT software used on the field.

For example, the personnel responsible for collecting blood samples should be familiarized with the part of the protocol that pertains to blood collection. The safety instructions for protecting the participant and the nurse or technician during the blood sample collection should be reviewed. Similarly those who will carry out the blood pressure measurements need specific information on why standardized blood pressure measurements are needed, what are the key steps in the measurement protocol, how the results are recorded and how the re-

sults are explained to the participants. The practical training will include e.g. carrying out adequate number of measurements observed by supervisors and feedback sessions. Detailed guidelines for the training and certification needed for each measurement will be provided in Part B of the EHES Manual.

15.2.2 Selection of the national trainers

The national trainers should:

- be well informed both on the aims and purposes of the national survey as well as on EHES standards, and
- have specific expertise in the subject area (e.g. survey ethics, blood pressure measurements).

The supervisors and persons with experiences from previous surveys can act as training assistants to train the other team members. They are needed in the practical training and in the role playing sessions.

15.2.3 Use of training materials and different training methods

The trainees should be encouraged to read the survey manuals before the training sessions, during and/or after the training. The survey manuals form the basis for all training. The EHES training materials will usually require national translations and adaptations. The training materials may include standard presentations, videos on interviewing and measurement techniques, and web-based education tools. Giving material to watch and read later at home and during the filedworkwill support learning. Newspaper articles and reports from previous surveys (if available) may help to see the importance of the survey and understand how the data will be utilized. An effective training programme will emphasize participatory exercises over lectures (Czaja & Blair 2004). If the fieldwork staff members do not practice their skills in a training session, they will practice them with real participants, which may lead to poor quality of data during the first days or even during the first weeks of the proper fieldwork. Role playing can be used in the participatory exercises, where the staff members take turns in playing different roles of the field work member (interviewer, measurer) and the participant.

In the role playing sessions those who play the role of the participants should be encouraged to vary their behaviour and to challenge the fieldwork member e.g. with asking several questions on the purpose of the survey, acting to be very busy, shy, fearful, reluctant or even aggressive. If feasible, practical training sessions can be recorded. Watching own performance helps to understand the purpose of standardization. Getting direct feedback during the practical sessions is important. Time needs to be allocated to discuss encountered difficulties and solutions directly after these exercises. The final step of the training should be to carry out the examination of an actual survey participant with supervisor observation.

Placing all training material and keeping a common discussion forum in the Internet (e.g. a specific survey training extranet site) will help to make sure that all staff members have up-to date information available throughout the fieldwork period. Open discussions between all field work members and other survey staff members should be encouraged during the training sessions. During the fielwork, meetings with the supervisors, audit visits and feedback sessions will support learning and remind the importance of standardization.

15.2.4 Duration and timing of the training

The EHES core measurements will require at least two or three training sessions, depending on the previous survey experience of the selected staff members. When the blood pressure is measured using the auscultation method, at least one week of training is required to ensure that all measurers have the same level. Each additional measurement will increase the duration of the training.

Training should be organized just before the fieldwork will be started. To allow substitution of other fieldwork team members when needed and rotating tasks (see Part A, Chapter 9) it is recommended that each team member will be trained to handle several measurements, even if the measurements are carried out by teams where the staff members have different tasks. Retraining during fieldwork should be organized if the fieldwork lasts for more than two or three months to ensure that the standards are kept. Retraining is essential also if observer effects or non-adherence to survey standards are observed during audit visits or by other forms of quality control during the fieldwork.

15.2.5 Certification

Certification for specific measurements is needed at least for the most challenging measurements requiring strict adherence to detailed protocols, such as blood pressure and waist circumference measurements and drawing blood samples. Certification is given after observed competent performance in practice and proven theoretical knowledge on measurement techniques and standardization.

References

- Czaja R, Blair J. Designing surveys. A guide to decisions and procedures. Sage Publications Ltd, Oine Firge Press, California, USA, 2004.

16. Preparation of the survey budget

Hanna Tolonen[1], Päivikki Koponen[1], Kari Kuulasmaa[1]

[1] National Institute for Health and Welfare (THL), Helsinki, Finland

PART A

Conducting a national health examination survey (HES) requires resources, which include personnel costs and materials as well as funds for travel, accommodation, rent, transport of materials, etc. The type and amount of resources needed depends strongly on the number of persons to be examined, the measurements to be done and the setting of the surveys. The preparation of the budget has to include the entire survey process (see Part A, Chapter 1) to ensure adequate resources for the planning and preparation, fieldwork as well as for the data analysis and reporting.

This chapter will provide guidelines for estimating the costs of the different phases of a national HES. It should be noted that these are just guidelines and have to be adjusted for the local situation. An Excel template (http://www.ehes.info/tc/tools/time_cost.xls), which may assist in preparation of the national HES budget, is also provided.

16.1 Purpose of the survey budget

The survey budget gives an estimate of the amount of money needed to carry out the planned survey components. With a well prepared survey budget, the work can be carried out without major surprises in the actual costs. The budget can also be used in discussions with the collaborators when possibilities to include additional measurements are negotiated. Adding a new measurement to the survey protocol will increase the total survey cost more than just the required equipment, trough longer examination times per person which affects the costs of survey personnel and survey site, and also through training, data management as well as data and material handling, quality control and reporting costs.

The funding available for the survey is always limited, and the survey budget has to be adjusted to the available funds. This may mean limiting the number of included measurements or the number of persons

to be examined from what was initially planned. Also the selection of the survey mode has to be considered in light of the funding available.

16.2 Components of the survey budget

For the national HES, 12 stages, which affect the survey budget can be identified:

1. Planning and preparation
2. Coordination
3. Sampling
4. Training of personnel
5. Dissemination (PR-activities)
6. Piloting
7. Recruitment of participants
8. Field work of the full-size HES
9. Laboratory analysis and sample storage
10. Data entry and cleaning
11. Quality assurance
12. Analysis and reporting

When the measurements to be included in the national HES have been selected and the survey setting has been decided, the time needed to examine one survey participant should be estimated. For example, let us assume a survey setting where the participants come to the fixed examination site and we measure height, weight, waist circumference, and blood pressure and draw blood samples for total and HDL cholesterol, and for fasting glucose measurements. Additional to that, the participants have to fill in the survey questionnaire at home, which is checked and completed at the examination site after the informed consent is explained to and signed by the participant. We can estimate that the checking of the questionnaire and obtaining informed consent will take 15 minutes, anthropometric measurements 10 minutes, blood pressure measurement 15 minutes and drawing the blood sample 15 minutes. These sums up to 55 minutes per participant.

The total time required to measure the entire sample can be calculated by multiplying the time per participant with the sample size. In practice, the number of survey participants to be measured will be less than the sample size. This should balance out the time needed for setting out the examinations sites, mandatory breaks of the fieldwork staff, re-training, etc.

For example, if we have a sample of 4000 persons and the time to measure one participant is 55 minutes, a total time required to measure the entire sample is 220,000 minutes = 3667 hours. If each field work day lasts 8 hours, this would mean 459 days. Depending on the number of parallel field work teams, the length of the field work period

can be calculated. If we have 4 field work teams, the field work would take 115 days and in case where the field work is conducted only during the working days from Monday to Friday, this would mean that the field work lasts 23 weeks. All this assumes that participants are examined at the fixed examination site one after other, without any overlap and without time needed for traveling (as in case of home visits). If the examinations can be organized with a field work team so that while one member of the team is examining one participant, the other one is at the same time examining the other one, i.e. there is overlap, the needed examination time decreases.

16.2.1 Planning and preparations

The planning process is described in Part A, Chapter 1. The main resource needed for the planning and preparation stage is personnel. The expertise of different professionals is needed for the planning. Each planning and preparation team should include or consult at least following experts:

- Survey leader, who has the main responsibility of the survey.
- Survey coordinator, who will organize the practicalities and will monitor the progress of the work.
- Senior researchers, who will provide epidemiological and public health perspective to the selection of the measurements and to the preparation of the survey questionnaires and manuals.
- ICT expert, who will plan and prepare the ICT infrastructure of the survey.
- Survey statistician, who will be consulted on sampling and analysis of the results.
- Press officer, who will be consulted on promotion of the survey as well as on dissemination of the results.
- Experts on the specific measurements, who will provide information on practical points of each measurement included in the survey.
- Laboratory experts, who will plan the collection of biological samples, sample processing, storage, transport and analysis.
- Expert on legal and ethical issues, who will be consulted in the questions relating to the data confidentiality, and ethical issues. This also includes obtaining the ethical approval for the survey, informed consent, etc.

There may also be need for a person with special knowledge on survey logistics. Survey logistics, scheduling of the examinations and transfer of personnel and materials can have a major impact on the survey budget. In some countries, translation of the survey questionnaires, etc. to different languages may be needed.

16.2.2 Coordination

The coordination activities of the survey are described in Part A, Chapter 1. The main resource needed for the coordination is the personnel but also some basic equipment and other resources are needed.

For the personnel, at least following is needed:

- Project leader, usually a senior researcher, who has the main responsibility for the survey.
- Survey coordinator, who will organize the practicalities and will monitor the progress of the work.
- Fieldwork supervisor, who will take care of the personnel management.

Often also an assistant(s) is(are) needed to assist with various practicalities, like recruitment of survey personnel, ordering the equipment and materials, mailing of the invitations, calling to the appointments for examinations, etc.

The coordination team (central office) needs at least computers with internet connection, telephones/mobile phones, printers, software licenses, and office materials. Also premises for the coordination office are needed, although they are often provided by the organizing institute. In many cases, the coordination team will also travel to the survey sites to promote the survey and to monitor the progress of the work. The travel and subsistence allowances need to be budgeted for this. Also some money usually has to be budgeted for the recruitment of field work personnel (newspaper advertisements, etc.).

16.2.3 Sampling

The definition of the sample size is described in Part A, Chapter 2 and the sampling process is described in Part A, Chapter 3. Depending on the local situation, the actual sampling can be done by a survey statistician hired to the survey team, or it can be bought as a service from a statistical institute or other sampling frame owner. In any case personnel and equipment are needed as well. Sampling has to be made for both the pilot survey and the full size survey.

Regardless of the way the actual sampling is done, at least following personnel is needed:

- Statistician to determine the adequate sample size, to sort out available sampling frame(s) and to design the sampling.
- Data manager to form a database where the sample information is stored.

Computers, and software licenses to establish a database for the sample are needed. Often the survey database is not established on a PC but on a server of the institute organizing the survey. Depending on the

institute, the use of server space and database platforms may or may not cost separately for the survey.

16.2.4 Training

The training programme is described in Part A, Chapter 15. Salaries and travel costs for the trainers and the field work staff to be trained, equipment and some other costs like preparation of the training materials have to be budgeted.

As trainers, at least following expertise is needed:

- Trainers for each measurement included in the survey. Depending on the qualifications of the trainers, one person can train the measurement protocols for several measurements or each measurement may need to be trained by different persons.
- ICT-support and/or data management person(s) to train how to use the ICT programmes designed for the fieldwork and how the data management is organized.
- Press officer to tell about the promotion activities for the survey.
- Statistician to tell about the sample selection as participants on the field may ask from the fieldwork personnel how just they got selected to this survey.
- Legal and ethical expert to explain the importance of data confidentiality and how to obtain the informed consent.

Equipment needed for the training depends on the included measurements (Part A, Chapter 5). A full set of equipment for each included measurement needs to be available during the training. A list of the required equipment for each recommended core measurement is given in Part B, Section 5.1 of the EHES Manual. It is good to provide each trainee with a folder, which includes training material together with the local survey manual.

In addition to the personnel and equipment, also premises for the training are needed. For the fieldwork staff to be trained and for the trainers, travel expenses and subsistence may also have to be paid.

16.2.5 Dissemination of information

The dissemination activities are described in Part A, Chapter 14 and Part C, Chapter 5. Dissemination of the national HES includes the establishment of the project image (brand), promotion of the survey before and during the field work as well as the dissemination of the survey results to the survey participants, various stakeholders, general public and the scientific community. The resources needed depends on the dissemination strategy of the specific survey, but some personnel resources and other costs should always be budgeted.

Regardless of the dissemination strategy, at least a press officer is needed to plan and supervise the survey dissemination strategy. There may also be need for a graphical designer to prepare the project image and promotional material for the survey. Often this service is bought from a service provider.

Depending on the planned promotion activities, there may be printing costs of promotional leaflets, advertisement costs for newspapers, radio and TV, and costs of press conferences and other promotion events. Costs from press conferences and promotion events may include payments to the publicly known persons who come to promote the survey, travel and subsistence costs, refreshments, etc.

If project web site is established, costs for setting up and maintaining the web site need to be budgeted. This may include personnel costs for web site designer, server and web domain costs as well as some royalty costs if for examples photos from photo banks are used for illustrations.

16.2.6 Piloting

The piloting process is described in Part A, Chapter 11. This requires both personnel resources, equipment and other resources.

For the personnel resources, at least following are needed:

- Full field work teams are needed to conduct examinations of the pilot sample.
- IT support to assist with the computer programs in use on the field.
- Laboratory personnel to process, transfer and storage of the blood samples and to carry out the laboratory analysis of the collected samples.
- Data manager to handle the incoming data.
- Statistician to assess and analyze the pilot survey data.

Examination equipment has been obtained for the training but additional sets may be needed for the pilot phase.

Other costs that may occur during the pilot are the printing costs of the invitation letters, informed consent forms and questionnaires, phone bills, rents of the examination sites, travel costs and subsistence of the field work and coordination personnel and transportation of the material to and from the field. Additional costs for the pilot phase come from the evaluation of the success of the pilot and from possible needs to change the survey protocol, to correct the computer programs, and to re-train the field work personnel.

Pilot survey can also be used to obtain photos for promotional material to be used in full-size survey. Therefore, costs for using a professional photographer may need to be included to the budget.

16.2.7 Recruitment of participants for the full-size HES

The recruitment process and ways to increase participation rate are described in Part A, Chapter 13. The used recruitment strategies are survey specific and need to be adjusted for the national situation.

For the recruitment of participants, at least following personnel is needed:

- Assistant, who will prepare the invitation letters and mail them out or in case the first or re-contact is through telephone, personnel who make the required phone calls.
- A designated person, who is named as a contact person and can provide more information about the survey when ever any person selected to the survey requires that.
- Data manager or assistant, who will make sure that the status of the recruitment and contacts, is also recorded to the survey database.

Computers and software licenses are needed to prepare invitations. It is highly recommended to establish a toll-free telephone number to which survey participants can call to change their examination times or ask additional questions. If a survey web site with for example web-based appointment scheduling system is established, personnel, equipment, and other costs relating to that have to be budgeted.

If recruitment also includes home visits to those not contacted by mail or telephone, salaries and travel costs for personnel doing the home visits has to be budgeted.

In some cases, incentives are provided for survey invitees. These can be cash, vouchers or small gifts. In case survey uses incentives, these need to be budgeted under recruitment activities.

16.2.8 Field work for the full-size HES

All participants of the survey will be examined during the field work. The selected survey setting (Part A, Chapter 7) will have a marked effect on budget. If questionnaires are filled in during the interview and are not self administered (Part A, Chapter 8), the cost of interviewers has to be included to the budget. When questionnaires are web-based or other data entry tools are used on the field, the programming of the web questionaires and data entry tools needs to be budgeted. Also in case where the examinations are done at the participant's home, the travel expenses for the field work staff have to be adjusted for this setting. It is good to remember that the coordination goes through out the survey process and is an essential part of the field work.

For the field work, at least the following personnel are needed:

- Full field work teams to carry out the measurements (Part A, Chapter 9.).
- ICT support to assist in the use of computer programs used in the field work (Part A, Chapter 12 and Part B, Chapter 6 of the EHES Manual).
- Data manager to handle the incoming data (Part A, Chapter 12).

There may also be need to have a medical doctor on call for consultation.

Most of the needed equipment (Part B, Chapter 5 of the EHES Manual) have already been obtained during the training and pilot phase but additional sets of equipment may be needed for the field work teams. It is good to have an extra set of equipment at the coordinating centre in case some equipment on field get broken or are not functioning properly.

Additional to the personnel and equipment, at least costs from printing of the invitations letters, informed consent forms, and quetionnaires, transport of the materials to and from the field examination site, transport of the personnel, accommodation and subsistence of the personnel during the field work, storage of the material in the field, rents of the examination sites, and data transfer, phone and internet connection have to be included in the budget, depending on the survey setting. The costs of the preparation and mailing of the feedback to the participants about their results (e.g. letters with laboratory results) has to be budgeted as well.

16.2.9 Laboratory analysis and sample storage for the full-size HES

The issues related to the laboratory analysis and sample storage are described in Part A, Chapter 10. Decisions on what is analyzed immediately and how much of the samples are stored for the future analysis effect the budget.

For the laboratory analysis, at least the following personnel are needed:

- Laboratory personnel with specific qualifications to handle the analysis.
- Data manager to handle the incoming data from the laboratory analysis.

A medical doctor should also be available for consultation if abnormal results are discovered.

Assuming that laboratory facilities are provided by the organizing institute, the following laboratory consumables need to be budged: aliquot tubes, pipettes, storage track and reagents. If the laboratory analysis is bought from the laboratory, these consumables should be part to the contract prize.

For the long time storage, the specific tubes which endure at least -70ºC, storage boxes and deep freezers of -70ºC are needed. For the deep freezers there has to be a security system to ensure that the samples do not melt during the power breaks and which will alarm if freezers get broken. Also log book of the stored samples is needed and has to be established if this is not already used in the laboratory.

16.2.10 Data entry and cleaning

The data management issues relating to the field work are described in Part B, Chapter 6 of the EHES Manual. The collected data have to be entered to the database unless already entered at the field using electronic/web questionnaires, and checked for completeness and correctness.

For data entry and cleaning, at least following personnel is needed:

- Data manager to maintain and update the survey database.
- Data entry persons in case data is entered manually.
- Statistician to work on data checking and cleaning.

Also transferring data forms to a data entry company and the data from there to the database will create costs. In case data forms are scanned/optically read, the scanning costs have to be included in the budget.

If computers and software licenses and server space needed for the maintenance of the database is not already budgeted during the planning phase, it should be included in budget here.

16.2.11 Quality control

The quality assurance of the survey process is described in Part A, Chapter 11 and the quality control of the data in Part C, Chapter 3 of the EHES Manual.

For quality control, at least following personnel are needed:

- Senior researcher/epidemiologist who know the survey protocol well and can detect deviations from it. Number and specific qualifications depend on the contents of the specific survey.
- Data manager who can detect possible systematic errors in the data already when they are entered to the database.
- Statistician to do basic checking of the data entered to the database.
- Laboratory personnel to perform both internal and external quality control of the laboratory analysis, and costs for taking part in the laboratory standardization program.

Equipment needed for the quality control of the measurements is specified together with the measurement procedures in Part B of this manual. In addition to these, computers and software are needed. An important part of the quality control is observing the work. There may be need to budget some travel and subsistence costs for the personnel conducting the quality control (audit visits) during the field work.

There may also be costs of the external quality control, such as for laboratory measurements.

16.2.12 Analysis and reporting

Issues relating to the analysis of the survey data and reporting are described in Part C, Chapters 6.

For the data analysis and reporting, at least the following personnel are needed:

- Data manager to maintain and share the survey data.
- Statistician to conduct the statistical analysis.
- Researchers to specify the research questions to be analyzed from the data and to interpret and report the results.

Additional to the personnel, also computers and software licenses are needed. In case the basic results are published in a form of book, the layout and printing costs need to be budgeted. Often results are also published on the web and a web manager and specific software may be needed to prepare the reports for the web.

16.3 Template for budget calculations

There is an Excel template at the EHES Web site under EHES Reference Centre Tools (http://www.ehes.info/rc/tools/time_cost.xls), which can be used while preparing the survey budget. The template is only a helping tool and each component has to be evaluated in the national situation. Instructions for using the Excel template are given in the first worksheet of the Excel template.