# European Data Watch

This section offers descriptions as well as discussions of data sources that are of interest to social scientists engaged in empirical research or teaching courses that include empirical investigations performed by students. The purpose is to describe the information in the data source, to give examples of questions tackled with the data and to tell how to access the data for research and teaching. We focus on data from German speaking countries that allow international comparative research. While most of the data are at the micro level (individuals, households, or firms), more aggregate data and meta data (for regions, industries, or nations) are included as well. Suggestions for data sources to be described in future columns (or comments on past columns) should be send to: Joachim Wagner, Leuphana University of Lueneburg, Institute of Economics, Campus 4.210, 21332 Lueneburg, Germany, or e-mailed to ⟨wagner@leuphana.de⟩. Past "European Data Watch" articles can be downloaded free of charge from the homepage of the German Council for Social and Economic Data (RatSWD) at: http://www.ratswd.de.

# National Health Data
# from the Robert Koch Institute

By Heribert Stolzenberg, Rüdiger Dölle and Bärbel-Maria Kurth

## 1. Health Surveys at the Robert Koch Institute

If you look up *Gesundheit* ("health") and *public use file* (PUF) in the (German) internet, you are most likely to find links to the corresponding websites of the Robert Koch Institute (RKI)[1]. A closer look will reveal that the data sources offered date as far back as the 1980s. Indeed the history of population-representative health surveys[2], on which the PUFs are based, began with the multi-centric German Cardiovascular Prevention Study (DHP) in April 1984. The three "waves" of what has become known as the National Examination

---

[1] See for example www.rki.de/nn_220830/puf.html.

[2] Then West Germany incl. West Berlin.

Surveys within the DHP today still represent the German reference source for health-related data on the period 1984–1991. The idea of offers a public use file was already discussed early on by members of the DHP research network. This idea of releasing data for secondary analyses seemed appropriate in view of the enormous cost of the study, which had been supported by substantial public funding. In consultation with the sponsors Federal Ministry for Research and the Federal Ministry of Health, and after an external expert appraisal of this project covering feasibility, data protection, scope of delivery and related overall conditions, a corresponding agreement was signed in the autumn of 1989. As a result the Institute for Social Medicine and Epidemiology of the former Federal Health Office (which became part of the Robert Koch Institute since 1994), was able to publish the first public use file on request that same year. Other representative health surveys followed.

The main purpose of the RKI health surveys is to generate a maximum amount of information on the state of health and health-related behaviour of Germany's resident population, while ensuring an optimum use of funds. The methodology – i.e. the sample design, the principles on operationalization and measurement, and data-collection techniques – is largely modelled on the tried-and-tested methods of empirical social research. The main topics of the Health surveys are summarized in Table 1.

*Table 1*

**Topics of the Health Surveys**

| Survey topics: |
| --- |
| Socio-demographics |
| Migration and origin |
| Disease process and symptoms |
| Consequences of disease, disability |
| Disease-related risk factors |
| Mental health |
| Psychosocial risk factors / pressures |
| Health-related quality of life |
| Utilization of healthcare services |
| Consumption of tobacco / alcohol / drugs |
| Physical / sporting activity |
| Accidents / injuries |
| Vaccinations |
| Health-related behaviour |
| Subjective health perception |
| Housing situation / environmental conditions |
| Leisure-time behaviour / media use |

*Health interview surveys (HIS)* use established survey techniques such as filling out questionnaires, computer-assisted telephone interviews (CATI), computer-assisted personal interviews (CAPI), and online polling via the internet or email.

However, the methodology also differs significant from purely sociological surveys, apart from the entirely different subject of "health" and the appropriate instruments needed for this purpose. It is the biomedical examinations, tests and medical-biochemical measurements that generate significant added value in addition to the results of the surveys; this part is referred to internationally as the *health examination survey (HES)*. Of course, the HIS/HES combination involves not only higher costs, but also a lot more logistical work. The total examination period per subject and the amount of time spent at each sample point must be carefully assessed, taking into account the required net sample and the planned duration of the study. Routes must be organized in such a way as to avoid, as far as possible, any regional or seasonal effects on the survey. If the examination and survey parts are well matched, this will definitely generate additional opportunities for validating individual results (see below).

Table 2 lists the health surveys that have been completed by the RKI to date or are currently ongoing. The upper section shows the surveys with the complex HIS/HES combination and field times of up to three years. The lower section lists the telephone health surveys (HIS type), whose survey period lasts an average of about eight months. Whereas the entire German adult population (with a landline telephone and sufficient knowledge of German) was included in the case of the CATI, for field access (HIS/HES) the sample was limited in the upper age range.

*Table 2*

**The RKI's Health Surveys**

| Name | Period | Net sample | Men | Women | Age group | Response rate |
|---|---|---|---|---|---|---|
| NUST0 (DHP) | 06/1984–04/1986 | 4,790 (west) | 2,417 | 2,373 | 25–69 | 66.7% |
| NUST1 (DHP) | 09/1987–10/1988 | 5,335 (west) | 2,649 | 2,686 | 25–69 | 71.4% |
| NUST2 (DHP) | 04/1990–05/1991 | 5,311 (west) | 2,623 | 2,688 | 25–69 | 69.0% |
| Survey Ost | 09/1991–06/1992 | 2,617 (east) | 1,223 | 1,394 | 18–79 | 70.2% |
| BGS'98 | 10/1997–03/1999 | 7,124 (total) | 3,450 | 3,674 | 18–79 | 61.4% |
| KiGGS | 05/2003–05/2006 | 17,641 (total) | 8,985 | 8,656 | 0–17 | 66.6% |
| DEGS | 11/2008–11/2011 | Approx. 7,800 (total) | | | 18–79+ | Approx. 52% |
| GSTel03 | 09/2002–03/2003 | 8,318 (total) | 3,872 | 4,446 | 18–80+ | 59.2% |
| GEDA09 | 07/2008–05/2009 | 21,262 (total) | 9,148 | 12,114 | 18–80+ | 51.2% |
| GEDA10 | 09/2009–07/2010 | 22,050 (total) | 9,567 | 12,483 | 18–80+ | 55.8% |
| KiGGS Wave 1 | 06/2009–06/2012 | Approx. 16,000 (total) | | | 0–24 | |

## 2. Health Monitoring at the RKI

Only insiders connect the name Robert Koch Institute with health surveys, especially since the latter rarely deal with infectious diseases. However, in line with its function as a public health institute, it is one of the main tasks of the Robert Koch Institute in Germany to recognize health-related problems in the general population in good time, and to observe and evaluate the development of these problems so as to be able to suggest measures to ensure and maintain the health of the population (Kurth et al., 2009). The RKI has been responsible for health reporting since 1999. In cooperation with the Federal Statistical Office, it compiles and repeatedly updates the necessary data basis from official statistical data, process data, and population registration data (http://www.gbe-bund.de). Appraisals of the available health data revealed gaps which could be repeatedly closed, albeit sporadically, by means of health surveys. As can be seen from Table 2, in the early days the surveys were not conducted regularly, nor were the intervals short enough to enable trends to be recognized in good time. External funds had to be raised to fund the individual surveys, making it impossible to plan properly. In recent years it has become increasingly apparent that surveys need to be repeated regularly to make it possible to observe health trends, evaluate health-policy measures and intervene to correct specific undesirable health developments. A concept for continuous health monitoring was therefore developed at the RKI (Kurth et al., 2005). This concept is based on three basic principles: use what exists, fill information gaps and interconnect data sources. Regularly recurring, methodologically comparable population-representative cross-sectional health surveys were planned. Since reliable (i.e. continuous) funding has been secured from the BMG for health monitoring, it has consisted of three components (Kurth et al., 2009) which dovetail closely in terms of both timing and content (see Figure 1).

As shown in Figure 1 and described in greater detail in Kurth et al., 2009, the RKI's health monitoring is made up of three components. One component is the German Health Update or GEDA (*GESUNDHEIT in Deutschland Aktuell*), a telephone cross-sectional survey which has already been conducted twice (RKI, 2011). Two elements were crucial when designing the content of GEDA: one was that it had to continue the health surveys conducted up to then; the other criterion was it should fit the other monitoring components as accurately as possible. GEDA is not only a new name, it also stands for a new orientation of telephone surveys at the RKI. This is reflected, for example, in the tripling of the net sample compared to the case numbers of its precursors. This means that regional analysis can now also be performed on the basis of the so-called Nielsen areas. Furthermore, evaluating this data set is made a lot easier by the creation and publication of a large number of standardized health-related indicators and indices. By including comparable variables from the GSTel03, the GEDA data can be used to analyse and evaluate over six years of health development in Germany (see on this RKI, 2009).

| KiGGS Studie zur Gesundheit von Kindern und Jugendlichen in Deutschland | DEGS Studie zur Gesundheit Erwachsener in Deutschland | GEDA GESUNDHEIT IN DEUTSCHLAND AKTUELL |
|---|---|---|
| Children and Adolescents | Adult | Adult |
| Long-term study (cohort) | Long-term study (Panel) | Cross-sectional study |
| Baseline Survey KiGGS 2003-06 HIS/HES | Baseline Survey BGS 1998 HIS/HES | Baseline Survey GEDA 2008-09 HIS (Telefonsurvey) |
| Wave 1 KiGGS 1 (2009-12) HIS (Telefonsurvey) | Wave 1 DEGS 1 (2008-11) HIS/HES | Annual repeat |
| **Continued funding through BMG and RKI** | | |

Figure 1: The three components of health monitoring

Another component is DEGS, a study on adult health in Germany. This latest health survey of the adult resident population in Germany was conducted between 2008 and 2011 in 180 locations in Germany and comprises a questionnaire part and an examination part (Projektbeschreibung, 2008). Because this survey again includes the subjects who took part in the BGS98, this is the first time that a longitudinal component has been integrated, making it possible to describe life histories and causal relationships. The first results are expected in the course of 2012.

The third monitoring component is KiGGS, a study on the health of children and adolescents in Germany. The entire group of children (and their parents) who took part in KiGGS from 2003 to 2006 have been interviewed again by telephone since 2009 (Projektbeschreibung, 2011). This survey will be completed in 2012, and another examination survey is already planned to follow in 2013. All the data collected in the context of health monitoring will be components of the Research Data Centre as public use files.

## 3. Examination Surveys

Apart from the fact that there is no central population register in Germany to date, the balance between logistical effort, costs and the demand for representatively collected data for nationwide health surveys can only be guaranteed if you have optimal and realistic sample planning. The population here consists of all the registered people living in private households in Germany during the survey and examination period; people living in military barracks, old people's homes, hospitals, and medical and nursing homes are excluded. The samples

are based on citizen registers; each is the result of a random selection stratified in two or three stages. The steps are as follows:

1. *First selection step: municipalities (sample points)* – The municipalities are stratified by state and municipality type (BIK classification) for selection. They are drawn with a probability that is proportional to the frequency of their size. The number of sample points to be drawn varies with the studies (see Table 4). Some major cities require two or more sample points per municipality. In addition to the presentation of national results, some surveys (e.g. BGS98) also call for separate representative data for eastern and western Germany. Accordingly, the points selected in these surveys are disproportional (EAST : WEST = 2 : 1).

2. *Second selection step: urban districts / constituencies* – In municipalities with between 50,000 and 100,000 inhabitants, one urban district is randomly selected within the municipality, which then represents the sample point. For municipalities with over 100,000 inhabitants the sample points correspond to randomly selected constituencies within the city limits.

3. *Third selection step: people* – The same number of people's addresses for the respective age groups are drawn in the selected sample points via the citizen registers.

The result of these selection steps is a gross sample that is representative of age, gender and municipality-size classes and is then made available to the field.

In the context of sampling it should also be mentioned that we know from other studies and our own pre-tests that some subpopulations (e.g. foreigners, migrants) are less willing to participate in the survey than the normal population. This effect is usually countered by slight over-sampling (factor 1.5 – 3), so that the affected results can be estimated more accurately and there are fewer missing values.

Each of the three health surveys within the DHP study consisted of a standardized examination and an extensive written questionnaire. These surveys focused on cardiovascular diseases and their risk factors. They also sought to supply information on the development of the state of health, health-related behaviour and health risks in Germany's adult population. After reunification, a health survey was conducted in eastern Germany in 1991/92 using largely the same methodology as the DHP surveys. The "Survey East" and the previously completed NUST2 provides a data basis that describes the status quo of health behaviour in the East and West German populations shortly after the German reunification.

The aim of the 1998 Federal Health Survey (BGS98) was to create and establish a routine set of instruments for public health reporting at the national level (Bellach, B.-M. et al., 1998). The BGS98 was the first all-German survey, i.e. it was carried out simultaneously in both eastern and western Germany accord-

ing to uniform characteristics and under the same conditions. By incorporating the data from the surveys conducted in 1990–1992 it is also possible to trace the development of the health differences that were determined between east and west in the early 1990s. The BGS98 survey programme consisted of a core survey and supplementary modules, most of which were conducted on subsamples of the study population; some modules were designed and co-financed by RKI cooperation partners (Gesundheitswesen, 1999).

Once a valid data basis on key health issues was available for the resident population of Germany aged between 18 and 79, the first National Health Interview and Examination Survey for Children and Adolescents (KiGGS) was launched in 2003 (Gesundheitswesen 2002; Kurth, B.-M. 2008). This survey, which was also modular in structure, was urgently needed, because up to then no comprehensive, nationwide health data on children and adolescents had been available in Germany. The logistics involved in the KiGGS study were significantly more complex again than in the adult surveys, since both the survey and examination parts had to be designed and applied in ways that were age-appropriate. The work of the field teams, too, was more difficult, since (small) children including carers were frequently present in the examination centre in addition to those who had been invited. Not only the young subjects, but also their parents had to be questioned in detail; in the case of children under 11 the essential information was provided only by the parents – via a questionnaire and a physician's interview (CAPI). Despite these more difficult conditions, KiGGS can be rated as very successful with a 66.6 % response rate. Taking the additional data from the shortened questionnaires for non-participants into consideration, basic information is available for almost 89 % of the adjusted gross sample. The fact that the KiGGS study is not only of public-health relevance, but also of importance for public-health policy – and equally serves epidemiological and medical research – is shown by the wide range of evaluation results that have been published to date (Bundesgesundheitsblatt, 2007; see also www.kiggs.de).

According to the plan, the fieldwork for the BGS98 successor DEGS are completed at the end of the year 2011. The release of the data as a PUF is scheduled for 2013.

Table 3 lists typical biomedical examinations and tests of the examination surveys. The main part of the surveys still consists of the written and/or oral interview (questionnaire and/or CAPI), in which the survey instruments (some of which are geared towards specific age groups) cover a wide range of topics (see Table 3 below). Some topics are covered by only a few individual questions, others by multi-column question groups. In order to be nationally and even internationally comparable with other surveys and studies, we try as far as possible to use validated standard instruments on the topics, or at least to largely gear them to proposed standards (e.g. for socio-demographics in Germany).

*Table 3*

**Examinations**

| Examinations and tests: |
| --- |
| Blood pressure / pulse |
| Anthropometry |
| Maturity status |
| Skin examination |
| Thyroid sonography |
| Cycle ergometry |
| Motor function tests (children) |
| Eye test |
| Blood / urine sample ($\rightarrow$ laboratory parameters) |
| Physical function tests (seniors) |

Finally, some examples of ways in which the HES data can be used will be shown: (a) The information provided by subjects on body size is incorrect in many cases. If such a subjective report is supplemented by a standardized measurement, the corresponding correction can be determined. Transferred to the survey level, it is ultimately possible to create a statistically valid correction model in this way to improve the corresponding HIS results (Kurth, B.-M./Ellert, U., 2010). (b) The current disease situation is often determined by a questionnaire which is answered by the subjects in the form of a self-report. The (subjective) disease prevalence that can be calculated from this information can be validated by other, more objective data and measurements in the case of combined HIS / HES surveys. For example, in the case of diabetes mellitus the subjects are asked in the physician's interview (CAPI) whether they have ever been diagnosed by a physician. By also asking the subjects what drugs they take, it is possible to check these against the drugs are usually prescribed for the condition. Corresponding figures for laboratory parameters such as glucose (blood / urine) and HbA1c (blood) can ultimately provide the decisive diagnostic information. Taking all these data together, the accuracy of self-reported data can be verified and a corrected disease prevalence determined in this way. Specifically the wealth of data measured and collected, and laboratory parameters analysed, make it possible to show topical reference percentiles and other distribution values for the population in Germany (Dortschy, R. et al., 2010; Neuhauser, H. et al., 2011).

## 4. Interview Surveys

In 2001 the RKI implemented an additional instrument for telephone interviews on health topics to provide policy-makers and scientists with topical health-related data in-between the health surveys with examination components,

but also for reasons of cost and efficiency. To ensure the continuity of the tried-and-tested theme-specific groups of questions, the electronic input forms for the first CATI at the RKI were programmed on the basis of the questionnaire and CAPI templates of the BGS98 survey. After a successful pre-test, the GSTel03 telephone health survey was finally launched in autumn 2002 (Kohler, M. et al., 2005). The sample is designed to reflect the adult, resident, German-speaking population living in private households who can be reached via landline telephone. The basis is a gross sample of randomly generated phone numbers drawn according to the Gabler-Häder method (Häder, S. / Gabler, S., 1998). Representativeness at the individual level is ensured by using the "next or last birthday method". The data collected in this way is weighted according to age group, gender distribution and region / federal state (*Land*), taking into account the number of people and telephone lines in the household and adjusted to the population structure in the survey year (see also v. d. Lippe et al., 2011; Schmich et al., 2012).

The relative speed with which such surveys can be completed compared to examination surveys, and the flexible way in which topical subjects can be integrated into the range of questions resulted in further telephone surveys like GEDA09 and the follow-up survey GEDA10, which will be released in early 2012 (RKI, 2011). In addition to the advantages of having an adequately staffed CATI, there are also other benefits: the "field work" can be better controlled; the survey period is relatively short; electronic data are directly generated; data-quality control is simplified; and the results can be made available quickly. The assessment is more ambivalent when it comes to access to the sample, the ease of reaching subjects, and the calculation of response rates. For example, the application of the above-mentioned Gabler-Häder method means there is practically no clustering, unlike in the case of the sample point design of the examination survey, which is based on data from the resident's registration office however, some population groups – such as immigrants, people with a low level of education, mobile-phone-only users, etc. – are greatly underrepresented. To reduce this selection bias, the marginal distributions of the three levels of education (data basis: Microcensus, Germany), which are based on the highest school-degree, are incorporated for the compilation of the post-stratification weight. At first sight, reaching a person in a telephone survey seems easier than attracting subjects for examination surveys with their limited survey periods per municipality. However, it depends a great deal on the planned number of attempts to call each person and on the amount of time the interviewing staffs have at their disposal. Although specifying the response rate makes it possible to give a differentiated assessment of survey quality, this only applies if this rate is clearly defined. In the absence of a German standard for this, we allocate the disposition codes from the CATIs to the AAPOR categories and then calculate the response rate on the basis of the different outcome rates.[3] In

---

[3] See the latest table at: www.aapor.org/Resources_for_Researchers.htm; the response rates reported for GSTel03 and GEDA are based on different outcome rates.

addition to the planned mean duration of a telephone interview, the quality of input-mask programming in particular is crucial for a successful CATI. A lot of experience and care is needed here, because faulty programming – especially in filter management – can have dramatic effects on participation (e.g. if subjects are irritated by redundant questions) and on the number of missing values (e.g. when areas are skipped). The skilful grouping of questions into memorizable individual questions, standard answers and standardized introductory and linking texts is also essential for telephone interviews.

## 5. Data Quality Assurance, Data Management and Documentation

The following section outlines how the IT-based data flow in our data centre currently functions – from data collection to the final PUF. The processing and quality assurance of the survey data is carried out in the epidemiological data centre in a process consisting of several steps and controlled by automated data workflow (see Figure 2). The aim of the data workflow is the standardized and automated processing of all necessary process steps for data quality assurance. As a rule the data run through several quality stages, beginning with the original data with unchanged content, progressing to raw data that have been adjusted for entry errors, followed by test data – perhaps containing individual corrections – and then to the audited final data. In the final data, all free texts are already largely automatically encoded and new variables created and filled with values. A subset of data is extracted for the public use file from this internal data set, which has been released for further statistical analysis. On principle, all necessary data changes – be they due to entry errors or filter violations – are only made indirectly using correction syntaxes and documented in a Quality-Assurance-database. This ensures that all data-set modifications can be traced and, if necessary, reversed at any time.

The workflow is controlled, and all survey items are managed, in a metadata directory known as Synopse-Database. Context information is stored and administered on every variable in this metadata directory. It includes hierarchical topic assignment, data type and data format, variable labels and value labels, as well as coding information. This metadata information is used in the process of data-set creation to automatically generate the final data sets for the desired statistical software systems from the audited data. In addition, the metadata directory enables users in the field of biometrics to synchronize and administer identical survey items across studies, and thus to generate harmonized longitudinal data sets.

The data documentation must be created parallel to the generation of the data set. It must both inform the PUF user about all important aspects of the relevant survey, and describe in detail the data processing and quality controls that

as been carried out. In addition to the labelled survey instruments and the code plans, the structure of the data set and its special characteristics must be documented in detail. In particular, newly formed variables (weighting factors, scores, indicators, indices) and their operationalizations must be explained. In addition, the documentation usually includes reports on results and methods and a list of bibliographical references to publications to date.
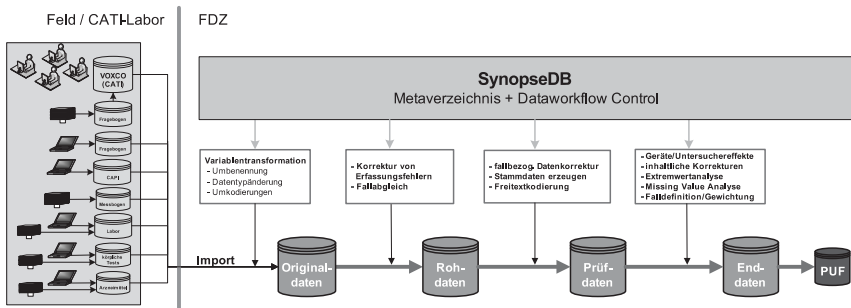


Figure 2: From data collection to the public use file

## 6. The Public Use Files

A public use file (PUF) is a file of research data infrastructure that is accessible for public purposes and can be evaluated by secondary analyses.[4] From the RKI's perspective, it is a data extract from a survey database which is made available on CD-ROM together with documentation files. As in the case of the social science archives, the data are offered in the commonly used formats of the major statistical programs SPSS, SAS, and STATA. Table 4 lists the currently available RKI public use files. The PUFs on the three DHP waves could originally be ordered individually, but this service has been discontinued in favour of an overall DHP file including documentation. The PUFs will appear (most likely) in a revised form in 2012 under the name DHPNUS.

RKI has been in charge of PUFs for more than 20 years. Questions asked by potential users typically tend to be of the following type:

1. How can I access the PUF data?

2. Where can I find out in advance whether any (relevant) questions / variables exist at all on a given evaluation subject?

3. Has the topic been processed or published (several times) before?

---

[4] The term scientific use file (SUF) is often used instead of PUF; some authors distinguish between the two on the basis of the extent and depth of the anonymization procedure used.

*Table 4*

**The RKI's Public Use Files**

| Survey name | Public use file | No. of variables | No. of sample points |
|---|---|---|---|
| NUST0 | } DHPNUS | 1,100 | 202 |
| NUST1 | | | 100 |
| NUST2 | } OW91* | 446 | 100 |
| Survey Ost | | | 50 |
| BGS'98 | BGS98 | 637 | 130 |
| KiGGS | KiGGS2003-06 | 1,397 | 167 |
| GSTel03 | GSTel03 | 193 | |
| GEDA09 | GEDA09 | 424 | |
| GEDA10 | GEDA10 | 402 | |

* Age range 25 – 69.

The answer to the first question leads to the RKI's homepage.[5] The corresponding application forms (2 pages and an information sheet) can be printed out directly via the menu items *Gesundheitsberichterstattung und Epidemiologie → Datenerhebungen → Public Use Files*. Please complete these application forms, sign and sent them by mail to the RKI address indicated. After your application has been checked and countersigned[6] by the project/departmental management, selected application data are entered in the PUF database and the PUF CD shipped to the address stated by you, along with a cover letter and the paying-in slip for the nominal fee (€90). This access method (which is relatively unbureaucratic for a higher federal authority) is based on the PUF regulations under the supervision of the BMG, which is also the data owner. The application forms are legally reviewed and updated when necessary by the RKI's legal department; the entire process is supervised from the beginning by our administration department.

Some survey variables are not released for the respective PUF, partly because of data-protection provisions and partly because of the need to provide[7] a PUF as quickly as possible after conclusion of the survey. Especially when new laboratory procedures, tests or complex survey instruments are used in a survey for the first time, it would be a risk to release such survey parts unchecked and without adequate documentation and instructions for evaluation. Nevertheless, PUF users can enter into cooperation agreements with the RKI enabling them

---

[5] Link: www.rki.de.

[6] Rare reasons for exceptions include information not provided in the application, etc.

[7] The period between the end of field research and release can vary considerably between one and three years.

to obtain free additional survey variables. The standardized procedure for this is described in the SOP (Standard Operation Procedure), which is also available from the RKI website on public use files. Once you have submitted (and substantiated) your cooperation application, it will be reviewed by an expert panel at the RKI. All documents are electronically recorded in a special database, which then controls the subsequent workflow. If the application is approved, and on receipt of the contracts signed by both institutions, the additional data are usually delivered immediately. Otherwise, a request is made for further information to help the panel decide on the application, or else the application is rejected, in which case reasons are given in writing. On the one hand, there is the "classic" form of cooperation in which both sides actively contribute on agreed topics and also publish jointly. On the other, there is the "passive" form of cooperation where the RKI only passes on the desired data and any related information to the applicant without claiming co-author rights. The RKI only needs to be consulted again prior to the submission of any results for publication to ensure there are no false or misleading interpretations of results.

Information on the health surveys can be accessed quickly via the corresponding RKI websites. In the case of some surveys it is also possible to view or download survey instruments in the form of PDF files. Up to now we have been happy to process and reply to all other requests direct by phone or email.[8] We plan to gradually build up a user portal to structure the available information in a more user-friendly fashion on the RKI's new website in future. The idea is to provide a collection of links to all survey instruments, supplementary and correction files on PUFs that have already been published, a list of FAQs and some evaluation tips. In addition to these IT solutions, the time-proven PUF telephone hotline, which has been used by many users in the past, will doubtless continue to play a role, time-intensive though it may be. The questions asked there tend to concentrate on methodology and advice on statistics or evaluation.

To avoid redundant evaluations of survey data wherever possible, the PUF application now includes the declaration that the user shall send the RKI either a copy or at least the corresponding bibliographical reference in the event of a publication based on the PUF data. Unfortunately, this does not always work in practice, so the RKI now sends the PUF buyers a corresponding reminder after a reasonable period of time. For its part the RKI compiles the lists of publications and publishes them in the internet. In future the RKI plans to compile a list of all bibliographical references based on a survey data set of which the RKI has knowledge. The planned portal solution could also further shorten the update cycles of these lists.

---

[8] Email: edz@rki.de; phone: 030/18754-3220.

## 7. PUF User Statistics

Having given an initial insight into the survey activities at the RKI, this section aims to shed some more light on the public use files and their users. In its regulations on PUFs, the BMG has stipulated that the RKI must be able to report on the PUF situation on request at any time. In the 1990s we therefore created a "PUF database" to store the most important information for such a report. This makes it easy to answer the question as to how frequently PUFs have been delivered to date (see Figure 3). The surprising thing has been the great demand for BGS98, which continues today – albeit significantly reduced. The PUF on the KiGGS study is also a "hit", while demand for the GEDA2009 data set published in February 2011 is only slowly growing. Its successor GEDA2010, to be published in early 2012, may attract more attention since it includes an additional weighting factor, making it possible to evaluate the comparable variables of both surveys together; as a result the number of cases available for analysis is approx. 42,000.
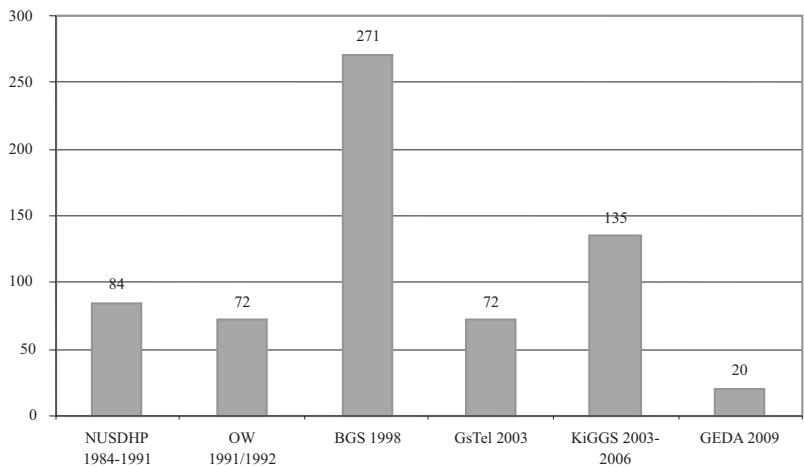


Figure 3: Number of RKI public use files delivered (last revised Oct. 2011)

When considering the number of deliveries listed, it is important to remember that hardly any of the recipients of the PUFs are individuals. In most cases these data sets are ordered by the managements or administrations of institutes. As expected, the main addressees are the faculties of university institutes in disciplines such as epidemiology, public health or medicine (see Figure 4). Because the data are representative, and many topics are not primarily assignable to the "health" category, these data are also attractive to sociological and economic institutes, among others. Possibly for lack of computer equipment,

health authorities and other healthcare players were hardly ever among the users at first, but this has changed significantly in recent years. Up to now the PUFs have rarely been ordered in non-German-speaking countries, so that, with a few exceptions, there has been no need to provide English-language versions of the PUFs. This may change over the next few years, however, as a result of European initiatives on health reporting.
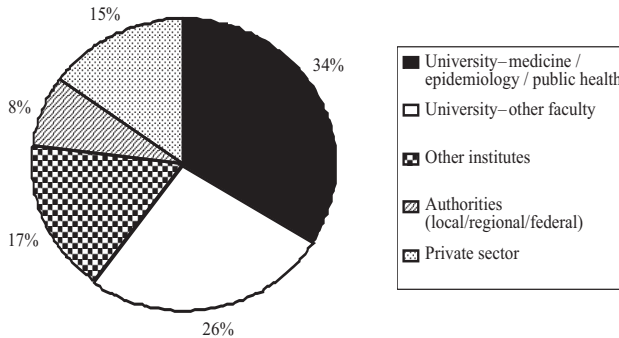


Figure 4: Institutions requesting public use files

The RKI is particularly interested in the reasons why orders are placed, i.e. what questions primarily triggered the user's interest in the respective PUFs. We use the word "primarily" here, because the users can, of course, carry out any analyses they wish in addition to the specified topics. We only hear about this in the event of a publication – assuming the user reports it to the RKI (in line with the contract). Figure 5 shows the distribution of the primary purposes in aggregated form; it reveals that the PUFs are used mainly as reference data sets when users are planning and/or assessing studies of their own. Although "specific questions" is the largest category, it is very heterogeneous and would be difficult to break down. The topics include complex statistical modelling, evaluations of variables outside the health sector, and cross-sectional issues. The aim of epidemiologically oriented institutes is usually to determine health-related risk and prevention factors and to discover possible impacting factors. Other users are interested in the morbidity situation in Germany and its correlations with socio-demographic information such as social status. The PUF data are often also used to analyse the extent to which people use medical services. Not least, PUFs are used for teaching purposes. The focus here is not only on the application of statistical methods using real data; the attached documentation is also used as a teaching aid. Despite the widespread use of PUFs and the large number of publications based on these data, we would, in conclusion, like to draw attention to the considerable remaining analytical potential of these data sets.
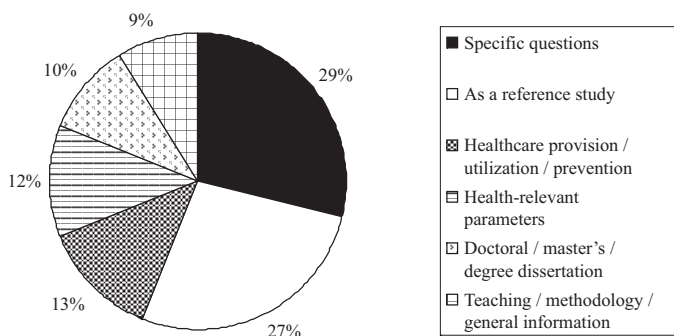
Figure 5: Primary purposes to which users put the public use file

## 8. Outlook

The RKI can look back on more than 25 years of experience with health surveys, so that in-house expertise is available for planning, data collection, data processing and the provision of quality-assured data sets of results. Of course, the RKI uses the survey data in many different ways to carry out its own tasks, as shown, for example, by the numerous published articles and booklets in the Federal Health Reporting field. In addition to the use of the data within the framework of the RKI's official duties, the importance of these data for epidemiological research should also be emphasized.

Following the accreditation of the RKI's Epidemiological Data Centre as a research data centre, we must extend both our online presentation of the data we make available to the scientific community and our description of its contents. A portal solution will soon be launched to give interested parties and users access to further documents and information in addition to the present project descriptions and options for ordering online.

The information provided will include:

– Lists of themes and variables available for interactive use,

– Descriptions of the instruments used,

– A list of FAQs on problems and instructions for use,

– Tips on evaluation including notes on statistical methodology,

– Survey-specific publication lists,

– A volume of tables.

An interactive user forum and online update service are also conceivable. One of our aims is a compact, user-friendly online presentation of the core results based on the PUF data in the form of tables, charts and/or thematic maps. In the medium term we intend to offer user further ways of accessing

the data in addition to the current method of shipping CD-ROMs. These will also include the possibility of teleprocessing, be it controlled (remote computing) or automated (remote data access).

# References

*Bellach,* B.-M. / *Knopf,* H. / *Thefeld,* W. (1998): Der Bundes-Gesundheitssurvey 1997/98 [The Federal Health Survey 1997/98]. Gesundheitswesen 60, Sonderheft 2, 59–68.

Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz (2007): Ergebnisse des Kinder- und Jugendgesundheitssurveys (KiGGS) [Results of the National Health Interview and Examination Survey for Children and Adolescents (KiGGS)]. 50 (Heft 5/6), 533–908.

*Dortschy,* R. / *Schaffrath-Rosario,* A. / *Scheidt-Nave,* C. / *Thierfelder,* W. / *Thamm,* M. / *Gutsche,* J. / *Markert,* A. (2009): Bevölkerungsbezogene Verteilungswerte ausgewählter Laborparameter aus der Studie zur Gesundheit von Kindern und Jugendlichen in Deutschland (KiGGS). Beiträge zur Gesundheitsberichterstattung des Bundes [Population-based distribution of selected laboratory parameters from the National Health Interview and Examination Survey for Children and Adolescents (KiGGS). Contributions on Federal Health Reporting]. RKI, Berlin.

*Forschungsverbund DHP* [DHP research network] (ed.) (1998): Die Deutsche Herz-Kreislauf-Präventionsstudie – Design und Ergebnisse [The German Cardiovascular Prevention Study – design and results]. Editors: von Troschke, J. / Klaes, L. / Maschewsky-Schneider, U. / Scheuermann, W., Bern / Göttingen / Toronto / Seattle.

Gesundheitswesen (1999) Schwerpunktheft: Bundes-Gesundheitssurvey 1998 – Erfahrungen, Ergebnisse, Perspektiven [Federal Health Survey 1998 – experience, findings, prospects]. Gesundheitswesen 61 (Sonderheft 2), 55–222.

Gesundheitswesen (2002) Schwerpunktheft: Kinder- und Jugendgesundheitssurvey – Konzept, Ziele, Aufbau, Kooperationspartner, Pretest [National Health Interview and Examination Survey for Children and Adolescents – concept, aims, structure, cooperation partners, pretest]. Gesundheitswesen 64 (Sonderheft 1), 1–130.

*Häder,* S. / *Gabler,* S. (1998): Ein neues Stichprobendesign für telefonische Umfragen in Deutschland [A new sample design for telephone surveys in Germany]. In: Gabler, S. / Häder, S. / Hoffmeyer-Zlotnik, J. H. P. (eds.), Telefonstichproben in Deutschland [Telephone survey samples in Germany], Opladen, 101–119.

*Kohler,* M. / *Rieck,* A. / *Borch,* S. / *Ziese,* T. (2005): Erster telefonischer Gesundheitssurvey des Robert Koch-Instituts – Methodische Beiträge. Beiträge zur Gesundheitsberichterstattung des Bundes [The Robert Koch Institute's first telephone health survey – methodological contributions. Contributions on Federal Health Reporting]. RKI, Berlin.

*Kurth,* B.-M. (2008): Gesundheitssurveys des Robert Koch-Instituts – Instrumente eines Monitorings der gesundheitlichen Versorgung der Bevölkerung [Health Surveys of the Robert Koch Institute – monitoring instruments for population health care]. In: Kurth, B.-M. (Hrsg.): Report Versorgungsforschung, Köln, 95–110.

*Kurth,* B.-M./*Ellert,* U. (2010) Estimated and Measured BMI and Self-Perceived Body Image of Adolescents in Germany: Part 1 – General Implications for Correcting Prevalence Estimations of Overweight and Obesity. Obesity Facts 3 (3), 181–190.

*Kurth,* B.-M./*Kamtsiuris,* P./*Hölling,* H./*Schlaud,* M./*Dölle,* R./*Ellert,* U./*Kahl,* H./ *Knopf,* H./*Lange,* M./*Mensink,* G. B. M./*Neuhauser,* H./*Schaffrath Rosario,* A./ *Scheidt-Nave,* C./*Schenk,* L./*Schlack,* R./*Stolzenberg,* H./*Thamm,* M./*Thierfelder,* W./*Wolf,* U. (2008): The challenge of comprehensively mapping children's health in a nationwide health survey: Design of the German KiGGS study. BMC Public Health 8, 196; doi:10.1186/1471-2458-8-196.

*Kurth,* B.-M./*Lange,* C./*Kamtsiuris,* P./*Hölling,* H. (2009): Gesundheitsmonitoring am Robert Koch-Institut. Sachstand und Perspektiven [Health monitoring at the Robert Koch Institute. Status and perspectives]. Bundesgesundheitsbl-Gesundheitsforsch-Gesundheitsschutz 52 (5), 557–570.

*Kurth,* B.-M./*Ziese,* T. (2005): Gesundheitssurveys des Robert Koch-Instituts. In: Streich, W./Braun, B./Helmert, U. (Hrsg.): Surveys im Gesundheitswesen – Entwicklungen und Perspektiven in der Versorgungsforschung und Politikberatung [Surveys in Health Care – trends and perspectives in health services research and policy advice]. St. Augustin, 15–22.

*von der Lippe,* E./*Schmich,* P./*Lange,* C. (2011): Advance letters as a way of reducing non-response in a National Health Telephone Survey: Differences between listed and unlisted numbers. Survey Research Methods 5 (3), 103–116.

*Neuhauser,* H./*Schienkiewitz,* A./*Schaffrath-Rosario,* A./*Dortschy,* R./*Kurth,* B. M./ *Ellert,* U./*Stolzenberg,* H. (2011): Referenzperzentile für anthropometrische Maßzahlen und Blutdruck aus der Studie zur Gesundheit von Kindern und Jugendlichen in Deutschland (KiGGS) 2003–2006. Beiträge zur Gesundheitsberichterstattung des Bundes [Reference percentiles for anthropometric measurements and blood pressure from the National Health Interview and Examination Survey for Children and Adolescents (KiGGS) 2003–2006. Contributions on Federal Health Reporting]. RKI, Berlin.

Projektbeschreibung DEGS (2009) Robert Koch-Institut (ed.): DEGS-Studie zur Gesundheit Erwachsener in Deutschland. Projektbeschreibung. Beiträge zur Gesundheitsberichterstattung. [DEGS – description of the projekt. Contributions on Federal Health Reporting ]. RKI, Berlin.

Projektbeschreibung KiGGS (2011) Robert Koch-Institut (ed.) KiGGS-Kinder- und Jugendgesundheitsstudie Well1. Projektbeschreibung. Beiträge zur Gesundheitsberichterstattung [KiGGS Wave 1 – Project Description. Contributions on Federal Health Reporting]. RKI, Berlin.

*Robert Koch-Institut* (ed.) (2009): 20 Jahre nach dem Fall der Mauer. Wie hat sich die Gesundheit in Deutschland entwickelt? Beiträge zur Gesundheitsberichterstattung des Bundes [20 Years after the fall of the Berlin Wall. How has health developed in Germany? Contributions on Federal Health Reporting]. RKI, Berlin.

*Robert Koch-Institut* (ed.) (2011): Daten und Fakten: Ergebnisse der Studie Gesundheit in Deutschland aktuell 2009. Beiträge zur Gesundheitsberichterstattung des Bundes [Data and facts: Results from the 2009 German Health Update (GEDA) study. Contributions on Federal Health Reporting]. RKI, Berlin.

*Schmich,* P./ *Jentsch,* F. (2012): Fieldwork Monitoring in Telephone Survey. In: Häder, S./ Häder, M./Kühne, M. (eds.), Telephone Surveys in Europe. Chapter 19. Berlin, 295–313.