DACSEIS

research paper series

No. 1

# The DACSEIS Project

Ralf Münnich, Rolf Wiegert

2001

## IST–2000–26057–DACSEIS

The DACSEIS research project is financially supported within the IST programme of the European Commission. Research activities take place in close collaboration with Eurostat.

`http://europa.eu.int/comm/eurostat/research/`

`http://www.cordis.lu/ist/`

`http://www.dacseis.de/`

---

# The DACSEIS Project

Ralf Münnich, Rolf Wiegert

Department of Statistics, Econometrics and Operations Research
University of Tübingen
Mohlstraße 36, 72074 Tübingen, Germany
e-mail: ralf.muennich@uni-tuebingen.de, rolf.wiegert@t-online.de

**Abstract:** Data quality has become increasingly important in social and economic statistics, especially when data are gained from various sources obtained using different methods. In many cases, measuring data quality is strongly connected to variance estimation as one error source. The DACSEIS project investigates variance estimation methods for complex surveys in the context of an applicable and comparable quality measurement of European household surveys.

The aim of the project is to deliver recommendations of how to measure data quality in complex surveys that take various aspects of error sources into consideration and allow the handling of different components of the measures. This will include a complex Monte Carlo simulation study in a practical environment, e. g. in synthetic, but realistic, universes that are close to selected European household surveys, to test the wide range of theoretical methodology *in practice*.

**Keywords:** Data Quality, variance estimation, complex surveys, nonresponse, imputation, Monte Carlo simulation

## 1 Introduction

*Data Quality in Complex Surveys within the New European Information Society* (DACSEIS), is a project within the *Information Society Technologies* (IST) programme of the European Commission, organised by the European Statistical Office *Eurostat*. The project started in March 2001 and will continue for three years. The main task of DACSEIS is to investigate data quality and reliability as well as estimating the precision of values gained from surveys with complex designs. Special emphasis will be put on European household and individual surveys such as the Labour Force Surveys or the Austrian and German Microcensus.

In the past years, precision of data and subsequently the measurement of data quality has become an increasingly challenging issue with the growing complexity of the surveys. This

problem has become even more acute due to the increased number of error sources that now have to be considered, especially in a European context where international comparisons need to be made. Within these sample surveys, many features and peculiarities, e. g. the complexity of the design, nonresponse rates and behaviour, have to be taken into account. Even more sophisticated is the measurement of the data quality in the context of investigating the different aspects of errors.

When dealing with data quality, many different facets can be considered. Following the Eurostat definition of quality in statistics, the following aspects play an important role (cf. Eurostat (2001a) or Eurostat (2001b)):

- Relevance of statistical concepts;
- Accuracy of estimates;
- Timeliness and punctuality of data dissemination;
- Accessability and clarity of information;
- Comparability of statistics;
- Coherence;
- Completeness.

The DACSEIS project basically concentrates on the accuracy of estimates and the comparability of statistics. The comparability of statistics, obviously, is an important task with regard to a harmonisation of data quality in Europe.

Within the European Statistical System (ESS), many different organisational forms exist in Official Statistics. Some countries have a central national statistical office where statistics is performed, covering data collection up to the dissemination of the results. Other countries have a centralised statistical system with several subunits. Within these possibly federal systems, generally, data collection is performed by these subunits, whereas the pooling of data and the dissemination of the results is centralised. In a further group of countries, statistics is performed within departments each concentrating on one or more statistical complexes, sometimes associated with different governmental departments. This heterogeneity of systems is intensified by different definitions of units as well as by various traditional and administrational aspects. Thus, the European Statistics System faces the difficult task of organising a harmonised and reliable European data base dealing with a heterogeneous environment. This induces a quite strong need for research developing methods and tools which allow for the preparation of data of comparable quality standards and valuable precision.

Although investigating data quality and its measurement has been of interest for a longer time, it has recently become increasingly important. Therefore, many studies and projects have already been setup, focusing on special aspects of data quality (e.g. SUP.COM 97-06, 97-14, and 98-16; c. f. `http://europa.eu.int/comm/eurostat/research/`). Unfortunately, little effort has been undertaken in finding adequate applications of the methods in complex surveys which calls for an exchange of technology from theory into practice. Thus, a very important issue is to investigate the above mentioned *accuracy of estimates*

in a practical environment, for instance simulations of populations of important applied surveys that allow for an easy application of more sophisticated research methods. This helps find suitable recommendations for the usage of these methods in applied surveys.

The importance of the assessment of quality can be drawn from the fact that all major statistical offices, e.g. the US Bureau of the Census, Statistics Canada, as well as Eurostat and its European member states, do research in measuring quality. Furthermore, Eurostat has its own *Working Group on Assessment of Quality in Statistics* and the *Task Force Variance Estimation*, which stresses the importance of further research in this field. DACSEIS particularly focuses on the research of variance estimation methods in complex surveys. The surveys of interest are mostly carried out by the National Statistical Institutes (NSI) to gain various types of data for a wide range of socio-economic problems.

# 2 Aim of the Project

The European Union in its present shape is an organisational and administrative unit still in the process of amalgamating. The European institutions, e. g. the European parliament, the board of ministries, the commission including Eurostat, are part of an approved organisation within the EU whereas their work aims to realise visionary targets as well.

A modern country with its administration and its economy will be positively influenced by and probably depend on an efficient organisation of official statistics. These official statistics have to preserve a reliable set of variables and data for diagnostic and analytical purposes to permanently get information about the recent economic and socio-demographic conditions in the Union. This is an important task for the EU in the context of its current status, but also in the context of a stepwise enlargement of the EU. Political development is currently devoted to a united, but federally organised European state and economy.

As mentioned above, European statistics face the difficult task of creating a harmonised and reliable socio-economic database for the economy in a united Europe with special emphasis on different national surveys and their international comparability. The definitions of units and variables used in the member states of the EU need to be standardised and the quality of the data gained from complex surveys e. g. household, population surveys and especially the labour force surveys should be made more homogeneous and comparable with respect to different quality components.

The core of the problem is to obtain applicable methods for variance estimation in complex multi-purpose sampling schemes. Simulated universes, reflecting the relevant national surveys and their complex characteristics and properties, will be generated. They are the basis of the analysis of the precision of the variance estimation methods regarding the influence on special designs and conditions of gaining data. A catalogue of recommended methods should enable the user to estimate variances effectively and reliably with comparable standards. A list providing criteria to be checked with certain complex design will facilitate the determination of a suitable variance estimation method for any specific estimation problem.

This usage of recommended methods can be a nucleus of a harmonised and standardised European quality management system in statistics. To fulfil this task, all relevant vari-

ance estimation methods currently available should be analysed, classified, evaluated and perhaps improved. This will be accomplished by theoretical research and by a realistic Monte Carlo study relevant for national surveys.

To speak generally, errors in data may have unfortunate consequences in economic and social analysis to an extent yet unknown. Therefore, research within DACSEIS is primarily a tool to advance the provision of reliable information on data quality and their dominating components.

The main goal of the project is to analyse the *accuracy of estimates* while taking into consideration different aspects of practical needs, like nonresponse rates and response behaviour, imputation, rotation schemes and applicability of methods in large scale universes. Additionally, an important task in this context is to develop efficient methods to combine data of surveys and registers. These methods are useful in reducing the response burden and may help improve the data quality, especially when dealing with rare events in small areas. A challenging issue in this case is the variance estimation.

The accuracy itself can obviously have many aspects that should be inspected. To find out the accurate procedure for measuring data quality means dealing with a large variety of different aspects which should be considered simultaneously. However, a univariate measure can obviously not sufficiently cover these different appearances of accuracy.

A thorough Monte Carlo simulation study is part of DACSEIS to allow for an investigation of survey procedures identical to those surveys which are practically applied in the team member countries. With respect to disclosure control, DACSEIS will use simulated universes for its research.

These considerations show the necessity of analysing the different components of accuracy and its relevant measures. Covering for instance estimators of variance, biases of the estimators, their mean squared errors, and complete simulated distributions of estimators as well as the influence of special properties of complex survey designs on the methods applied. In this context, the influence of nonresponse and imputation on accuracy is also an important goal.

A final result of the work will be the dissemination of the results of the research by a so called *recommended practice manual* that will be built up for all potential users as a database.

# 3 The Research Team and Partners

The research team consists of four university and four official statistics institutions. The institutions and their corresponding key persons are

**Centraal Bureau voor de Statistiek (CBS),** The Netherlands
Paul Knottnerus (PKTS@cbs.nl)

**Eberhard Karls University of Tübingen (UT),** Germany
Ralf Münnich (Co-ordinator; ralf.muennich@uni-tuebingen.de)

**Johannes Kepler University Linz (JKU/IFAS),** Austria
Andreas Quatember (Andreas.Quatember@jk.uni-linz.ac.at)

**Statistisches Bundesamt (StBA),** Germany
Wolf Bihler (Wolf.Bihler@destatis.de)

**Swiss Federal Institute of Technology (EPFL),** Switzerland
Anthony Davison (Anthony.Davison@epfl.ch)

**Swiss Federal Statistical Office (SFSO),** Switzerland
Jean-Pierre Renfer (Jean-Pierre.Renfer@bfs.admin.ch)

**Tilastokeskus (Statistics Finland; StatFi),** Finland
Seppo Laaksonen (Seppo.Laaksonen@stat.fi)

**University of Southampton (SOTON),** United Kingdom
Chris Skinner (cjs@socsci.soton.ac.uk)

Directly involved in the DACSEIS project is also Alois Haslinger, Statistics Austria. Additionally, Jan Bjørnstad, Statistics Norway, as well as Statistics Sweden, are permanently co-operating with the DACSEIS research team. The internal evaluator of the project is Siegfried Gabler, ZUMA Mannheim, Germany (gabler@zuma-mannheim.de).

Two other FP5 projects are closely related to some workpackages that are also part of DACSEIS (cf. section 4). To ensure a good interaction among these projects, DACSEIS was integrated into a cluster with EUREDIT, a project on data editing and imputation (co-ordinator: John Charlton, ONS; `http://www.cs.york.ac.uk/euredit/`) and EURAREA, which is on small area estimation techniques (co-ordinator: Patrick Heady, ONS; `http://www.statistics.gov.uk/methods_quality/eurarea/`). Further co-operation is in progress with the quality group of the *amrads* project (`http://amrads.jrc.it/`).

# 4 DACSEIS Workplan

The research of the project is organised in 11 workpackages (WP). As shown in figure 1, the investigation of quality issues as described earlier in the paper, is done in WP1 *Variance estimation in complex surveys*. This workpackage also includes a summary of the results of the other workpackages in the form of *best practice recommendations*. To accomplish this challenging objective, different surveys will have to be investigated with respect to the underlying universes, their corresponding sample survey, as well as their peculiarities, e. g. nonresponse rates and behaviour, weighting schemes etc. The surveys to be examined within WP2 *Structures and analysis of relevant national surveys* basically cover household surveys. They are a very important input for WP3 *Monte Carlo simulation study of European surveys*, where the estimation procedures that will be investigated within the methodological WPs are applied to practical situations. These practical situations will be simulated from the surveys mentioned above as well as from artificial surveys to find out the behaviour of the variance estimators within a *synthetic, but realistic* environment. The conclusions from this Monte Carlo simulation study will be summarised for the *recommended practice manual*.

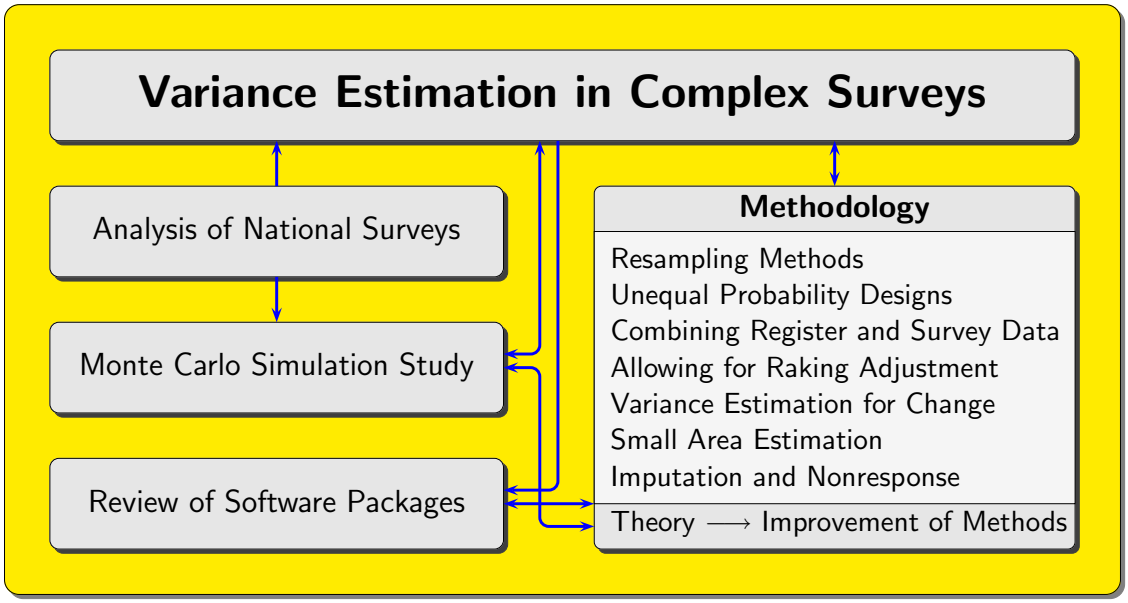


Figure 1: Outline of the structure of the workpackages within the DACSEIS project

The methodology mainly focuses on variances estimation of sampling errors for cross-sectional and to some extent also for longitudinal data. The methodological framework of variance estimation techniques can be seen from figure 1 and will be described in more detail below. Additional work will be done by the examination of variance estimation methods for non-sampling errors with respect to missing values and their imputation.

Finally, one workpackage will focus on the applicability of software packages with special emphasis on measuring data quality.

A detailed overview to the working programme is as follows (WP leader in brackets):

**WP1: Variance estimation in complex surveys (UT)**

The aim of this workpackage is to summarise and conclude the results from the methodological workpackages. The methods will cover variances estimation methods as well as attempts to deal with more complex error measures, e. g. the mean squared error or different variance components. A comparison of the different methods and approaches will help to achieve valuable results in terms of an adequate applicability of the methods. The comparison will be supported by a Monte Carlo simulation study based on synthetic but realistic universes (c. f. WP3).

To be able to evaluate the methods of interest, a catalogue of criteria has to be set up. The presentation and the evaluation of the methods of interest will conclude in a *recommended practice manual*. This manual will be provided as an electronic database to facilitate the handling of the methods of interest including experience on their applicability with different survey designs.

**WP2: Structures and analysis of relevant national surveys (JKU/IFAS)**

As an important step in the examination of the variance estimation methods, different surveys of the participating countries of the project, primarily household surveys, have to be studied with respect to the structure of the underlying universe,

the arrangement of the universe as sampling units as well as the sampling procedure itself. The surveys to be examined are the British, Dutch and Finish labour force surveys (LFS), the Austrian and German microcensus, and the German and Swiss household budget Survey (EVS/HBS).

In addition to the survey and sampling procedures, further peculiarities of the surveys, e. g. nonresponse rates and behaviour, will be considered to allow for a more detailed understanding of the survey and its features as well as for a more realistic set-up of the simulation study.

**WP3: Monte Carlo simulation study of European surveys (UT)**

One major contribution to the project regarding the applicability of the methods to be examined will be the Monte Carlo simulation study. The aim of the study is to examine the variance estimation methods in the context of *close to reality surveys* and of artificial surveys that are of special interest for the practical implementation of the methodology of variance estimation techniques.

To achieve a more valuable examination of the methodology, the set-up of the study is very important. A flow chart as an overview of the simulation set-up is depicted in figure 2 (for a detailed overview of the notation see `http://www.dacseis.de/` → `research`).

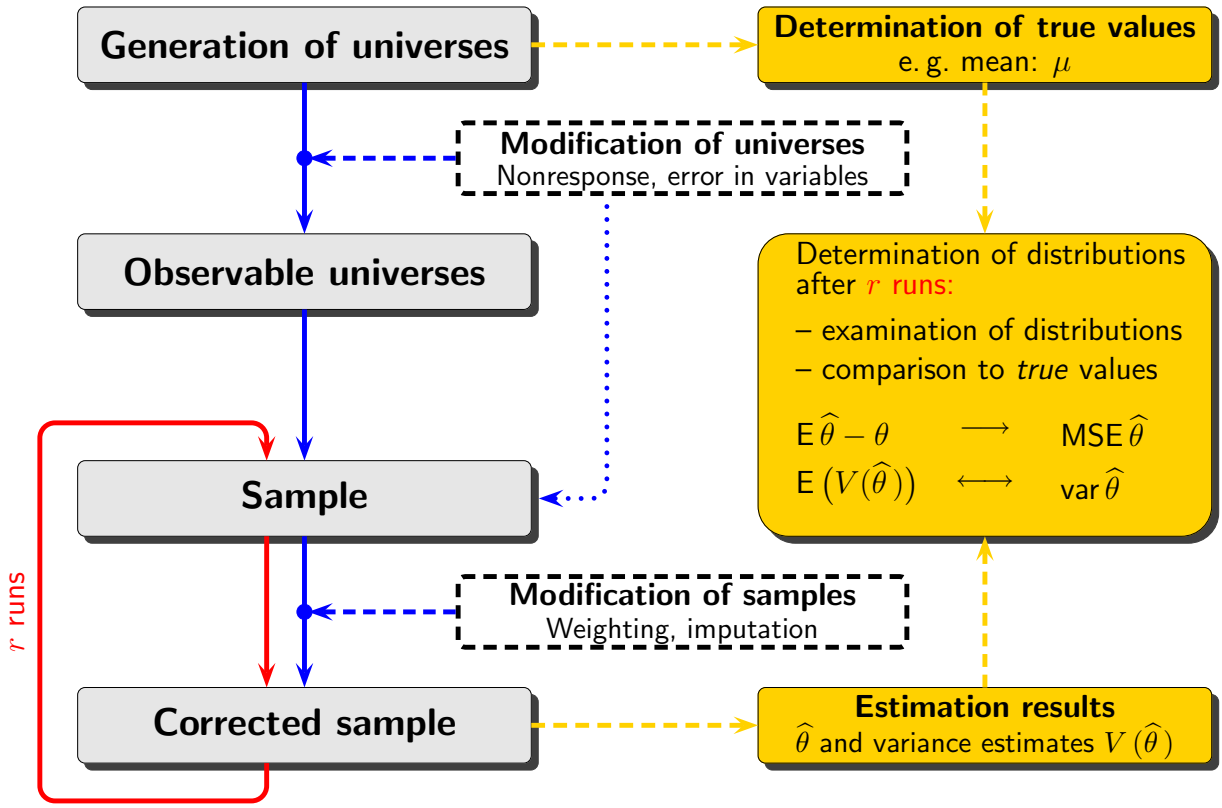


Figure 2: Outline of the simulation study

The first step of the simulation study is to generate *close to reality* universes from the national surveys presented in WP2. These universes have to be generated to be

able to easily identify the sampling units including the relevant information from the variables of interest.

For a later comparison of the estimation and variance estimation methods among themselves and with the *true* values, the true values have to be determined from the *true* universe. If needed, the universe has to be modified to an observable universe taking into regard the peculiarities of the national surveys mentioned above. This will make it possible to introduce nonresponse to the study by identifying nonrespondent units within the universe. At this stage, other effects, e. g. errors in variables, can be added to a model-based view. The flexibility of the model allows also for introducing special effects directly into the sample, e. g. special types of nonresponse that are connected to the surveying process.

To gain estimated distributions for the estimators and variance estimators, many recurrences of the sampling procedure have to be performed. Within each recurrence estimates of the values of interest and of the variance estimates of the estimator are computed. After $r$ runs, which in general varies between 10000 and 100000 depending on the problem, the simulated distributions will be examined and compared to the true values. This approach allows to examine the variance estimators themselves as well as the different components of the variance, e. g. the sampling variance or the variance due to nonresponse.

The Monte Carlo simulation study will be implemented in C++ under LinuX to guarantee a fast, flexible and effective computation. The computational effort can vary enormously depending on the application of sophisticated imputation techniques for the modification of the samples, on the variance estimation techniques in which expensive resampling techniques may be used and, obviously, on the runs of the study itself.

**WP4: Review of software packages for variance estimation (StatFi)**

The aim of this workpackage is to review software packages currently in use, especially at National Statistical Institutes, with respect to their capabilities in the field of *variance estimation in complex surveys*. The software packages to be reviewed will cover general statistical software packages, e. g. SAS, SPSS, S-Plus and R, as well as specialised statistical software packages, e. g. Bascula, SUDAAN, WesVar, CLAN and Poulpe.

The results of this workpackage should enable the end-user to make his decision regarding his own needs. Therefore, various aspects of the evaluation will be considered, such as computational aspects, methods implemented, programming facilities, usability, handling of complex data and much more. The comparison of the software packages will also include a test on real datasets to evaluate the performance under realistic conditions.

**WP5: Resampling methods for variance estimation (EPFL)**

When applying modern sample designs, complex estimation procedures have to be taken into account. As a rule, reliable mathematical analysis of these procedures is difficult and in many cases impossible. Instead of highly sophisticated analytical methods - which only work well for specific problems - resampling methods have

been advocated to compute variances for survey estimates. Due to their computational burden, resampling methods have not been widely used in European practical statistics. Additionally, only little attention has been paid to their practical implementation in complex surveys.

The objective of this work is a systematic evaluation and comparison of resampling methods for variance estimation in complex surveys, allowing for post-stratification, seasonal adjustment, panel structure, nonresponse and other features of real data. This should result in concrete and implementable recommendations for adoption by national and EU agencies in various national and pan-national surveys.

Evaluation of the practical performance of resampling methods is another target of these investigations.

**WP6: Variance estimation for unequal probability designs (Soton)**

Application of unequal probability designs faces the difficult task of determining joint inclusion probabilities. Generally, this task is of great complexity.

To estimate variances in the presence of unequal probability design, sampling textbooks often refer to the Horvitz-Thompson and Yates-Grundy variance estimators. In practice, however, these variance estimators are only applicable in special problems in which it is possible to calculate positive inclusion probabilities.

Therefore, simpler variance estimators will be investigated within this WP. They depend only on first order inclusion probabilities. Linearisation techniques and jackknife estimation will be applied to maintain methodological results which work without the use of complex calculations and joint inclusion probabilities.

Simulation studies are planned which will be evaluated with regards to the performance of those methods. Results will enhance the practical tools for estimating reliable variances.

**WP7: Variance estimation in the case of combining register and survey data (CBS)**

This workpackage focuses on an adequate use of register data for statistical purposes. In the EU, a large number of registers in administration and in official statistics already exist and are possibly ready for exploitation in the EU countries. They can, for instance, be found in labour, financial or business administration. These registers should be investigated and classified regarding their usefulness especially for statistical purposes. However, difficulties will appear while dealing with disclosure control, with different definitions of variables of administration and of statistics and last but not least with a proper matching of records from different sources.

Firstly, relevant registers and their data currently used in European statistics will have to be inspected. This should be done by means of a poll to get a thorough overview of the European register landscape. Then different European registers may be identified for further statistical use.

Secondly, algorithms for the combination of data via advanced statistical methods should be improved and applied. The currently existing methods seem to pave the way for better punctuality of the data and their dissemination while saving costs in the data generating process without loss of data quality. Response burden can

be reduced too. Furthermore, an evaluation of these methods should be performed with special emphasis on data quality. A test based on the data of the German Microcensus will help investigate these methods and their performances.

In the Netherlands, a new approach towards coherence of statistical output will be investigated. It deals with the so-called micro database and reference database which are in use with this concept. The micro database is the result of matching separate data sources (survey data and administrative register data) of the same object-type. Based on the variables in the micro database the reference database contains accurate and mutually consistent estimates of finite population parameters. This database is set up by successively adding recent estimates taking into consideration earlier estimates.

The estimation technique to accomplish this is based on repeatedly applied regressions. The CBS team of DACSEIS intends to develop a strategy to estimate the corresponding variances and to evaluate the performance of this approach.

Thirdly, case studies relating to simulations should be carried out to gain reliable results.

**WP8: Allowing for raking adjustment (Soton)**

The estimates gained in many household surveys in Europe are adjusted to different socio-demographic variables which are real indicators of the status of the households. They are a kind of complementary information that should be taken into account. The adjustment will be carried out by repeated application of post stratification with respect to the different socio-demographic variables.

Alternative approaches will be taken into consideration to allow for such raking adjustment. Effects of linearisation and application of jackknife techniques for variance estimation in this context should be investigated. One of these promising approaches is to approximate the adjustment by applying regression methods. This is another kind of weighting procedure.

A problem with this procedure may be the large number of covariates in the regressions. Their number can be very large e. g. several hundred. This can lead to some problems with the computation. To avoid these obstacles, alternative approaches will be considered and evaluated.

**WP9: Variance estimation for change (Soton)**

The methods described above are generally devoted to cross-sectional surveys or at least to estimates at certain time points. When turning to longitudinal data, changes of a variable may become of interest, e. g. the change in the number of unemployed or in the unemployment rates. As time is varying, many aspects of the survey may vary as well and lead to interactions between the effects induced by the time change and the estimation process itself. These changes can appear in longitudinal designs and in rotational designs.

Therefore, one may have to deal with different response patterns and thus with varying weighting schemes. This workpackage will investigate alternative approaches to estimating variances of change. The impact of temporal stratification will also be considered, in particular the allocation of the sample to different weeks within the

year will be considered. The study will take into consideration some of the surveys that are included within WP2.

**WP10: Variance estimation for small area estimates (UT)**

In many surveys, estimation on the basis of subpopulations is of interest. These estimations can be regarded as rare events even in complex surveys. In general, direct estimates of the variables of interest may then lead to unnecessarily high variances of the sampling error.

The aim of small area estimation methods is to borrow strength from additional information from other sources. Thus, reliable information on the variable of interest can be obtained without directly applying the sample information on the variable of interest within the small area or domain. Unfortunately, the efficiency of these methods is strongly influenced by the interaction between the estimation and the auxiliary variables. Additionally, severe biases may occur which causes the consideration of other error measures than the variance which in this case may end up in misleading estimation results. This workpackage will primarily focus on the German data situation and the data quality measurement in the context of small area estimation.

**WP11: Imputation and nonresponse (StatFi)**

If data are missing, e. g. due to nonresponse, there are several groups of methods which estimate variances.

Newer methods to substitute missing data are based on multiple imputation. Multiple imputation methods have already been tested in bio-statistical and biometric applications. In this workpackage, prototype software like Solas and Mice will be examined to evaluate their performance in this field. Recent multiple imputation techniques are not automatically applicable to National Statistical Institutes. Some of the possible advantages of the multiple imputation methods should be tested closer and modified for common use in those cases where MI techniques are discovered to be advantageous. Nevertheless, single imputation techniques will also be investigated because the presumption is that single imputation techniques will be superior to multiple methods, at least when providing completed micro data files for standard end-users.

In addition to standard variance estimation methods in this field, resampling techniques will be tested for their applicability to imputation methods. Currently, there are many gaps in applying resampling techniques because to some extent they depend strictly on the imputation methods applied. If the multiple imputation-influence is not transparent, it is not possible to judge the quality of variance estimation. This dependency should be examined thoroughly.

The National Statistical Institutes often use specially tailored procedures if they apply proper variance estimation methods. First of all, more general solutions are required. For this purpose, the WP contributes in two directions. First, a kind of good practice manual will be written, its focus being on the estimation of imputation variance. Second, attempts to provide some new research using empirical complex survey data will be done. Reports and computer codes of successful attempts will be available. WP11 will pay attention to the achievements of the EUREDIT project.

Statistics Finland, University of Southampton and Statistics Netherland are involved in this project with the examination of multiple imputation methods.

Further workpackages will cover the variety of organisational aspects as well as the completion of the final report and the dissemination of all expected results.

# 5 Expected Achievements of the Project

The project should achieve substantial compilation, classification and sufficient evaluation of the relevant variance estimation methods in complex survey designs. A broad documentation for the users will summarise the results.

The simulation of relevant socio-economic universes with emphasis on realistic Monte Carlo studies will enhance the know-how of this type of semi-theoretical analysis remarkably. All relevant procedures of estimating variances adequately will be contained in an intended *recommended practice manual* and be evaluated by Monte Carlo Simulations.

In addition, the theoretical research for the improvement of estimation procedures will serve the practically oriented research of applied simulations. All the results of this careful and broad research will be collected and set up as a manual of recommended practice for the benefit of researchers as well as end-users. Source and pseudo code of the procedures involved will also be demonstrated and disseminated. Last but not least, the project work of DACSEIS is carried out to obtain standards of measuring the error influences consistently and to obtain a comparative overview over the data quality of some of the most important household and individual surveys in the National Statistical Institutes within the EU. These experiences will be a basis for an implementation of a harmonised and standardised data-platform within the European statistics.

The following lists the main DACSEIS goals:

- Substantial compilation, classification and evaluation of relevant methods for variance estimation in complex surveys.
- Simulation of relevant socio-economic universes by Monte Carlo studies adjusted to reality.
- Theoretical research to improve of estimation procedures.
- Monte Carlo simulation to evaluate the procedures and their performances.
- Cataloguing of user-oriented instructions and criteria for variance estimation methods in practice.
- Thorough documentation of results in a *recommended practice manual* in form of a database.
- Documentation of source and pseudo code of relevant variance estimation methods

# 6 Exchange of Ideas with the End-User Group

An important aspect of the development of the *best practice* recommendations for variance estimation in complex surveys is to meet end-user demands. Therefore, it is important to ensure communication between the end-user group and the research team. Two ways of exchange have already been put into operation, the Eurostat expert groups on quality in statistics and the amrads project (`http://amrads.jrc.it/`).

A direct exchange of ideas between the end-user group and the DACSEIS research team is planned for Summer 2002. Until then, the DACSEIS *recommended practice manual* is planned to be implemented as an electronic database which will work as a knowledge database and a platform to allow for communication.

The exchange of ideas has particular importance within WP2. Likewise, this exchange can serve as an input for WP3 to enrich the practically oriented simulation study with more details on features of the national surveys. To facilitate this flow of information, a questionnaire on national individual and household surveys and on variance estimation methods in use was sent to the end-user group via the Eurostat expert group on assessment of quality in statistics.

# 7 Final Remarks

The DACSEIS project aims to study data quality measurement methods that basically focus on variances estimation of sampling errors, with special emphasis on the applicability of the methods. This will be done by giving support to the theoretical research on a variety of variance estimation techniques and on variance components themselves by a large Monte Carlo simulation study. This study is based on various surveys from different countries. The methodology, even if concentrated on household surveys, should also make it possible to apply the techniques in a bigger context of sampling surveys.

The intention is to support the link between theoretical research and the dissemination of sophisticated models to applied and Official Statistics by an electronic *recommended practice manual* as a knowledge database. This challenging approach will require the exchange with potential end-users to reach its full functionality in the context of a best practice manual.

DACSEIS is organised as an international research project with the above mentioned partners. The whole work is co-ordinated by close contacts, exchange and discussion on various platforms. A steering committee frequently meets to control and evaluate the results in the course of the DACSEIS project. The quality of the results will be evaluated firstly by the internal evaluator and secondly the external evaluator, who is appointed by Eurostat.

The DACSEIS research project started in March 2001 and will continue until February 2004. The dissemination of the final results will cover the traditional ways of publication in journals and as a final report as well as in electronic-based media as stated for the *DACSEIS recommended practice manual of variance estimation in complex surveys*. Additionally, a final conference is planned for 2004.

More details on the project can be found in MÜNNICH (2001), its progress and activities can be found at the DACSEIS homepage

`http://www.dacseis.de`

# Acknowledgements

The DACSEIS project is financially supported within the *Information Society Technologies* programme (IST-26057-2000) of the European Commission (`http://www.cordis.lu/ist/`). The project was placed for the cross programme action CPA8 which is coordinated by Eurostat (c. f. `http://europa.eu.int/comm/eurostat/research/`).

The authors want to thank Jean-Louis Mercy, Timo Alanko and Christiana Letteri, Eurostat, for the support within the contract negotiation phase and the project so far, as well as Harald Sonnberger and Håkan Linden, Eurostat, for the co-operation within the WG on quality in statistics and the task force variance estimation.

Finally, this project is obviously not only the work of the coordinator, but was developed with the contributions of all partners. Therefore, the authors want to thank the DACSEIS research team for the very good co-operation, which actually started in October 1999 with the preparation of the proposal.

# References

**Eurostat (2001a)**: Working Group on Assessment of Quality in Statistics (Doc. Eurostat/A4/Quality/00/General/Definition).

**Eurostat (2001b)**: Working Group on Assessment of Quality in Statistics (Doc. Eurostat/A4/Quality/00/General/Standard).

**Münnich, Ralf (2001)**: Data Quality of Complex Surveys within the New European Information Society (DACSEIS). In: Proceedings of the 2001 ETK/NTTS Conference in Crete.