Marianne Haapea

NON-RESPONSE AND INFORMATION BIAS IN POPULATION-BASED PSYCHIATRIC RESEARCH

THE NORTHERN FINLAND 1966 BIRTH COHORT STUDY
MARIANNE HAAPEA

NON-RESPONSE AND INFORMATION BIAS IN POPULATION-BASED PSYCHIATRIC RESEARCH
The Northern Finland 1966 Birth Cohort study

Academic dissertation to be presented with the assent of the Faculty of Medicine of the University of Oulu for public defence in Auditorium 1 of the Department of Psychiatry, on 23 April 2010, at 12 noon
Haaapea, Marianne, Non-response and information bias in population-based psychiatric research. The Northern Finland 1966 Birth Cohort study
Faculty of Medicine, Institute of Clinical Medicine, Department of Psychiatry, Institute of Diagnostics, Department of Diagnostic Radiology, Institute of Health Sciences, University of Oulu, P.O.Box 5000, FI-90014 University of Oulu, Finland
Acta Univ. Oul. D 1049, 2010
Oulu, Finland

Abstract

Study samples in medical research are selected according to the objectives of the studies. Researchers seek to collect data as extensively and reliably as possible. In practice, however, data are often missing or may be incorrect. This thesis covers some of the problems concerning missing data and data collection in psychiatric research. Methods for adjusting for missing data and for evaluating the reliability of data are presented.

The data originate from the Northern Finland 1966 Birth Cohort (N = 12058). This study explored how participation in an epidemiologic study that includes questionnaires and a clinical examination is affected by mental health (N = 11540), and whether non-participants experience more severe clinical symptoms than participants in a psychiatric field study (N = 145) among subjects with a psychosis. Inverse probability weighting (IPW) was used to adjust for non-participation in comparisons of brain volumes between schizophrenia and control groups. The precision of self-reported medication use was also explored (N = 7625).

In an epidemiologic study of all cohort members, subjects with a psychiatric disorder participated less actively than those without one. In the psychiatric field study, non participants were more often patients with schizophrenia than other psychoses. The psychiatric symptoms of non-participants were more severe and they needed more hospital care than participants. The use of IPW led to higher estimates of cerebrospinal fluid volume and lower estimates of grey and white matter volumes in schizophrenia patients, and increased the statistical significance of the differences in brain volume estimates between the schizophrenia and control groups. The precision of self-reported data on psychoactive medication use was substantial.

Due to non-participation, the true prevalence of psychiatric disorders is probably higher than the prevalence estimates from field studies that are based on data provided by participants only. In order to reflect the true differences in the target population, weighting methods can be used to improve estimates affected by non-participation. Regarding psychoactive medication use, data collected by postal questionnaire can be assumed accurate enough for study purposes. However, it may underestimate the prevalence of medication use due to non-participation.

Keywords: Birth cohort, follow-up, missing data, non-participation, non-response bias, precision, register, reliability, schizophrenia, self-report, weighting
Tiivistelmä

Tutkimusaineisto valitaan tutkimuksen tavoitteiden perusteella. Tavoitteena on kerätä kattava ja virheetön aineisto. Käytännössä kuitenkin osa tietoja voi puuttua tai olla virheellistä. Tässä väitöskirjassa esitellään yleisesti menetelmiä huomioida puuttuva tieto analyyseissä ja arvioida aineistojen luotettavuutta psykiatrisessa tutkimuksessa.


Epidemilogisessa tutkimuksessa ne kohortin jäsenet, joilla oli jokin psykiatrinen sairaus, osallistuivat passiivisemmin kuin ne, joilla ei ollut psykiatrista sairautta. Psykoosipotilaat, jotka eivät osallistuneet psykiatriiseen kenttätutkimukseen, sairastivat tutkimukseen osallistuneita useammin skitsofreniaa kuin muita psykooseja ja heidän taudinkuvansa oli vakavampi. Painottamien kasvattavat aivovolyymien ja alensi harmaan ja valeen aineen tilavuuksien skitsofreniapotilailla, ja lisäsi aivovolyymien erojen tilastollista merkitsevyyttä skitsofreniapotilaiden ja vertailuhenkilöiden välillä. Itse ilmoitetun psykoaktiivisten lääkkeiden käyttötiedon luotettavuus oli merkittävä.

Kadosta johtuen psykiatristen sairauksien todellinen vallitsevuus on todennäköisesti korkeampi kuin vallitsevuuden estimaatit, jotka on laskettu tutkimukseen osallistuneiden tiedoista. Painotusmenetelmä voidaan käyttää parantamaan puuttuvan tiedon vääristämää estimaatteja, koska painottamalla huomioidaan todellisia eroja kohdeväestössä. Tutkimuksessa luotettavuus postikyselyllä kerätyin aineiston voidaan olettaa olevan laadultaan riittävä tutkimustarpeisiin.

Asiasanat: kyselytutkimus, osallistumattomuusharha, painottaminen, puuttuva tieto, rekisteritutkimus, seuranta, skitsofrenia, syntymäkohortti, tarkkuus, toistettavuus
Acknowledgements

This work was carried out at the Departments of Psychiatry and Diagnostic Radiology, University of Oulu, in collaboration with the Institute of Health Sciences, University of Oulu. This dissertation study is a part of the Northern Finland 1966 Birth Cohort study.

I wish to express my warmest gratitude to my supervisors for guiding me throughout this work. I am thankful to Adjunct Professor Jouko Miettunen, Department of Psychiatry, University of Oulu, who has given numerous hours of his time to guide my work, help me with practical issues, and build up my spirits. I am most grateful for the patience he has generously offered. I am grateful to Professor Matti Isohanni, Department of Psychiatry, University of Oulu, for giving me the opportunity to carry out this study as a member of his research group. I appreciate his advice, ideas, and the financial support. I am thankful to Academy Research Fellow Juha Veijola, Academy of Finland, for his advice and support. His optimism has been invaluable.

I am sincerely grateful to Professor Kristian Wahlbeck, National Institute of Health and Welfare, Helsinki, and Adjunct Professor Jari Haukka, National Institute of Health and Welfare, Helsinki, for their useful comments, constructive criticism and advice as the reviewers of this thesis.

I wish to express my gratitude to Professor Osmo Tervonen, Head of the Department of Diagnostic Radiology, Oulu University Hospital, for facilities and support during this work, and for the opportunity to work as a member of inspiring research groups.

The co-authors of the original articles deserve my sincere gratitude. I had the honour of working with a group of specialists and inspiring people, including Professor Matti Joukamaa, Tampere School of Public Health, University of Tampere, and Department of Psychiatry, Tampere University Hospital; Professor Hannu Koponen, Department of Psychiatry, University of Kuopio; Professor Esa Lääärä, Department of Mathematical Sciences, University of Oulu; Professor Marjo-Riitta Järvelin, Department of Epidemiology and Public Health, Imperial College, London, UK; MD, PhD Erika Jääskeläinen, Department of Psychiatry, University of Oulu; MD Päivikki Tanskanen, Department of Diagnostic Radiology, Oulu University Hospital; and MD, PhD Sari Lindeman, Department of Psychiatry, University of Oulu. I would like to thank them for their efforts in the original articles, and their encouragement throughout this work.
I am grateful to Professor (Emerita) Paula Rantakallio, who planned and collected the original cohort data in the 1960s. Professor Rantakallio made it possible for me and other researchers to use this unique data.

I wish to thank the staff in the Departments of Psychiatry and Diagnostic Radiology for the friendly working environment during the years of this work. I want to especially thank Ms Minna Lakkapää and Ms Pirkko Kaan, Department of Psychiatry, University of Oulu, for all their practical help during this work. I am also thankful to Ms Tuula Ylitalo, Department of Public Health Science and General Practice, University of Oulu, for her practical help. I wish to warmly thank authorized translator Keith Kosola for the careful revision of the language of this thesis.

The following Foundations and Pharmaceutical companies are acknowledged for their financial support: the Academy of Finland, the Sigrid Jusélius Foundation, the Stanley Medical Research Institute, the European Commission, NARSAD, the Jalmari and Rauha Ahokas Foundation, the Lundbeck Foundation, the Research Foundation of Orion Corporation, the Health Care Foundation of Northern Finland, and the Oulu University Pharmacy Foundation.

Sincere thanks to my dear parents, Ritva and Juho Kokkoneni, for their love, understanding, and financial and mental support. Thank you for helping in the practical issues of everyday life in many ways.

Finally, I want to give my special and deepest thanks to my dearest sons Miska, Mikko and Miikka, and my husband, Mika Haapea. Thank you.

Oulu, March 2010

Marianne Haapea
## Main concepts

<table>
<thead>
<tr>
<th><strong>Main concepts</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>Agreement with gold standard</td>
</tr>
<tr>
<td>Attrition</td>
<td>A special case of wave non-response when a subject does not return to participate in a study</td>
</tr>
<tr>
<td>Auxiliary variables</td>
<td>Study variables that are not of interest themselves but rather background information on all eligible subjects</td>
</tr>
<tr>
<td>Available-case analysis</td>
<td>Only observations with missing variable values are deleted, i.e. sets of units with the parameters observed are used in statistical analyses (also called pairwise deletion)</td>
</tr>
<tr>
<td>Bias</td>
<td>Statistical sampling or testing error caused by systematically favoring some outcomes over others</td>
</tr>
<tr>
<td>Bias index</td>
<td>Measure of effect of bias</td>
</tr>
<tr>
<td>Cohen’s kappa</td>
<td>Measure of agreement of categorical items</td>
</tr>
<tr>
<td>Complete-case analysis</td>
<td>Only observations for which all values are observed are used in statistical analyses (also called listwise deletion)</td>
</tr>
<tr>
<td>Eligible</td>
<td>Subjects who fill the inclusion criteria</td>
</tr>
<tr>
<td>Exclusion criteria</td>
<td>Criteria for subjects who are not to be sampled</td>
</tr>
<tr>
<td>Inclusion criteria</td>
<td>Criteria for subjects who are to be sampled</td>
</tr>
<tr>
<td>Ineligible</td>
<td>Subjects who do not fill the inclusion criteria</td>
</tr>
<tr>
<td>Inverse probability weighting</td>
<td>Method to compute weights to adjust for non-participation, based on propensity to participate</td>
</tr>
<tr>
<td>Item non-response</td>
<td>Partial data of a subject is available with missing values on certain individual items</td>
</tr>
<tr>
<td>Loss to follow-up</td>
<td>A subject cannot be traced to be invited to participate</td>
</tr>
<tr>
<td>Missing data mechanisms</td>
<td>Types of missingness</td>
</tr>
<tr>
<td>Missing</td>
<td>Probability that a subject remains in study depends on exposure or confounders, but not on the outcome</td>
</tr>
<tr>
<td>Missing at random</td>
<td>Probability that a subject remains in study does not depend on exposure, confounders, or the outcome</td>
</tr>
<tr>
<td>Not at random</td>
<td>Probability of dropping out depends on the outcome and cannot be completely explained by covariates</td>
</tr>
<tr>
<td>Missingness</td>
<td>Any data missing from data</td>
</tr>
<tr>
<td>Multiple imputation</td>
<td>Observed values are used to generate a range of plausible values for each missing value</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Non-participation</td>
<td>Subject intended to be sampled does not participate in a study (also called non-response)</td>
</tr>
<tr>
<td>Non-response</td>
<td>Failure to obtain a measurement on study variables</td>
</tr>
<tr>
<td>Non-response bias</td>
<td>Bias caused by participants differing from non-participants with respect to outcome variable</td>
</tr>
<tr>
<td>Participation rate</td>
<td>Response rate</td>
</tr>
<tr>
<td>Patterns of missingness</td>
<td></td>
</tr>
<tr>
<td>Arbitrary</td>
<td>Any set of variables may be missing for any subject</td>
</tr>
<tr>
<td>Univariate</td>
<td>Certain items are completely observed, but some other items are not</td>
</tr>
<tr>
<td>Monotone</td>
<td>Certain items are completely observed. When item is missing for a subject, all items following the missing item are also missing</td>
</tr>
<tr>
<td>Population</td>
<td>Entire group of individuals that information is collected from</td>
</tr>
<tr>
<td>Precision</td>
<td>Reliability when comparing two data of which neither is gold standard</td>
</tr>
<tr>
<td>Prevalence index</td>
<td>Measure of effect of prevalence</td>
</tr>
<tr>
<td>Prevalence-adjusted and bias-adjusted kappa</td>
<td>Version of kappa that adjusts for prevalence and bias</td>
</tr>
<tr>
<td>Propensity score</td>
<td>Scored propensity to participate in a study</td>
</tr>
<tr>
<td>Proportion of negative agreement</td>
<td>Proportion of units with agreement on the values indicating absence of outcome in two data sets</td>
</tr>
<tr>
<td>overall agreement</td>
<td>Proportion of units for which two data sets agree</td>
</tr>
<tr>
<td>positive agreement</td>
<td>Proportion of units with agreement on the values indicating presence of outcome in two data sets</td>
</tr>
<tr>
<td>Reliability</td>
<td>Consistency of a set of measurements</td>
</tr>
<tr>
<td>Response rate</td>
<td>Proportion of responders (participants) in a sample (also called participation rate)</td>
</tr>
<tr>
<td>Sample</td>
<td>Part of population examined in order to gather information</td>
</tr>
<tr>
<td>Single imputation</td>
<td>Missing item is replaced by its approximate</td>
</tr>
<tr>
<td>Unit non-response</td>
<td>Subject’s follow-up data is completely missing</td>
</tr>
<tr>
<td>Validity</td>
<td>Degree to which an instrument measures what it is set out to measure</td>
</tr>
<tr>
<td>Wave non-response</td>
<td>A subject is present for some waves of data collection and missing for others in longitudinal studies</td>
</tr>
<tr>
<td>Weight adjustment</td>
<td>Complete cases are weighted in order to make data representative of the full sample or population with respect to auxiliary variables</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Weighting class method</td>
<td>Method to compute weights, based on homogeneous (with respect to auxiliary variables) classes of eligible subjects</td>
</tr>
</tbody>
</table>
**Abbreviations**

- **ATC**: Anatomical Therapeutic Chemical classification system
- **ANOVA**: Analysis of variance
- **BI**: Bias index
- **CSF**: Cerebrospinal fluid
- **DDD**: Daily defined dose
- **DSM**: Diagnostic and Statistical Manual for Mental Disorders
- **ECA**: Epidemiologic Catchment Area study
- **FHDR**: Finnish Hospital Discharge Register
- **GM**: Grey matter
- **ICD**: International Statistical Classification of Diseases and Related Health Problems
- **ICV**: Intracranial volume
- **IPW**: Inverse probability weighting
- **κ**: Cohen’s kappa
- **MAR**: Missing at random
- **MCAR**: Missing completely at random
- **MI**: Multiple imputation
- **MNAR**: Missing not at random
- **MRI**: Magnetic resonance imaging
- **NEMESIS**: Netherlands Mental Health Survey and Incidence Study
- **NFBC 1966**: Northern Finland 1966 Birth Cohort
- **OR**: Odds ratio
- **PABAK**: Prevalence-adjusted and bias-adjusted kappa
- **PART Study**: Population study of mental disorders in Stockholm
- **p_e**: Proportion of expected agreements
- **PI**: Prevalence index
- **p_neg**: Proportion of negative agreement
- **p_o**: Proportion of observed agreements
- **p_pos**: Proportion of positive agreement
- **RR**: Relative risk
- **WM**: White matter
List of original publications

This thesis is based on the following articles, which are referred to in the text by their Roman numerals.


Contents

Abstract
Tiivistelmä
Acknowledgements 7
Main concepts 9
Abbreviations 13
List of original publications 15
Contents 17
1 Introduction 21
2 Review of the literature 23
   2.1 Missing data ......................................................................................... 23
      2.1.1 Sampling .................................................................................... 23
      2.1.2 Types and patterns of missing values ........................................... 24
      2.1.3 Factors associated with non-participation .................................... 26
      2.1.4 Retaining response rates in studies .............................................. 32
      2.1.5 Reporting of missing data ............................................................. 33
   2.2 Statistical methods in handling missing data ........................................ 34
      2.2.1 Procedures based on completely recorded units ........................... 34
      2.2.2 Single imputation ....................................................................... 35
      2.2.3 Multiple imputation .................................................................... 36
      2.2.4 Weighting methods ..................................................................... 37
   2.3 Precision between self-reported and pharmacy data on medication use
      2.3.1 Self-reported data on medication use ........................................... 39
      2.3.2 Register data on medication use ................................................... 41
   2.4 Agreement statistics between two data sets ........................................... 43
      2.4.1 Proportions of agreement ............................................................. 43
      2.4.2 Cohen’s kappa .............................................................................. 45
      2.4.3 Effect of prevalence and bias ........................................................ 46
      2.4.4 Prevalence-adjusted and bias-adjusted kappa ................................ 47
3 Aims of the study 49
4 Material and methods 51
   4.1 Study population: The Northern Finland 1966 Birth Cohort .............. 51
   4.2 Epidemiologic health study (31-year follow-up) ................................... 51
   4.3 Psychiatric field study .......................................................................... 54
      4.3.1 Invitation and participation of subjects with a psychosis ............... 55
4.3.2 Invitation and participation of control subjects ............................ 55
4.4 Data on medication use in the NFBC 1966 ..................................... 57
4.5 Collection of data in psychiatric studies of the NFBC 1966 ............ 58
  4.5.1 Data on psychiatric morbidity .................................................. 59
  4.5.2 Other data used in this thesis .................................................... 60
  4.5.3 Variables used in the original studies ...................................... 60
4.6 Statistical methods ....................................................................... 63
4.7 Ethical considerations and personal involvement ........................... 66

5 Results

5.1 Original study I ............................................................................... 67
  5.1.1 Effect of psychiatric morbidity on participation rates ............... 67
  5.1.2 Effect of gender and education on participation rates ............. 67
  5.1.3 Effect of the data collection method on participation rates ...... 69
5.2 Original study II ............................................................................. 70
  5.2.1 Non-participation among subjects with a psychosis .............. 70
  5.2.2 Non-participation among control subjects ............................. 72
5.3 Original study III ............................................................................ 73
  5.3.1 Non-participation among subjects with schizophrenia .......... 74
  5.3.2 Brain volume estimates in subjects with schizophrenia ......... 74
  5.3.3 Comparing weighted and unweighted analyses of brain
       volumes between the schizophrenia and control groups .......... 75
5.4 Original study IV ............................................................................ 77
  5.4.1 Prevalence estimates of medication use according to
       pharmacy and self-reported data .............................................. 77
  5.4.2 Precision of self-reported medication use ............................... 78
  5.4.3 Effect of socio-economic factors on the precision of self-
       reported medication use ...................................................... 79
  5.4.4 Effect of non-participation on self-reported medication
       use .................................................................................. 79

6 Discussion

6.1 Main findings .................................................................................. 81
6.2 Discussion on the results ............................................................... 82
  6.2.1 Original study I ................................................................. 82
  6.2.2 Original study II ............................................................... 84
  6.2.3 Original study III ............................................................... 87
  6.2.4 Original study IV ............................................................... 88
6.3 Methodological considerations .................................................... 89
6.3.1 Participation in the studies of the NFBC 1966 ......................... 89
6.3.2 Statistical methods for dealing with non-participation .......... 90
6.3.3 Reliability of self-reported data ............................................. 92
6.4 Strengths and limitations .......................................................... 93
  6.4.1 Original studies I-II .............................................................. 93
  6.4.2 Original study III ................................................................. 94
  6.4.3 Original study IV ................................................................. 95

7 Conclusions ................................................................................. 97
  7.1 Main conclusions ................................................................. 97
  7.2 Future research .................................................................. 98

References ...................................................................................... 101

Original publications .................................................................... 111
1 Introduction

Study samples in medical research are selected according to the objectives of the studies. In the ideal situation in research, all subjects would participate with a response rate of 100%. However, this is hardly ever the real situation. Data are often missing or may be incorrect, which may affect the generalisability of the results. In longitudinal and cohort studies, attrition tends to increase during follow-up, and may be significant in population-based studies, especially when studying psychiatric morbidity.

Missing data reduce sample size and statistical power, and may introduce non-response bias. They may affect the differences in outcome variables between cases and controls, as those with certain risk factors may cumulate into non-participants or participants.

Attrition with respect to psychiatric disorders has been studied in, e.g. the following epidemiological studies: Epidemiologic Catchment Area study (Eaton et al. 1992, Badawi et al. 1999, Eaton et al. 2007), Netherlands Mental Health Survey and Incidence Study (Bijl et al. 1998, de Graaf et al. 2000), general population study of mental disorders in Stockholm (Lundberg et al. 2005), Prenatal Determinants of Schizophrenia Study (Susser et al. 2000), and Finnish longitudinal survey on depression (Eerola et al. 2005).

Participation in epidemiological studies has been associated with socioeconomic, demographic and health-related factors, and with participation in earlier studies. According to Drivsholm et al. (2006), subjects who participate in health studies have better health profiles than those who do not participate. Some previous studies on psychiatric patients have suggested that more psychiatric problems occur among non-participants than among participants (Fischer et al. 2001, Vanable et al. 2002).

Not only non-participation, but also the quality of collected data affects the generalisability of the results. Many population studies are based on data collected by postal questionnaire. However, the reliability and validity of data obtained with self-report inquiries on medication use depend on the ability and will of the subjects to reply accurately (e.g. Caskie & Willis 2004), as well as on data collection methods and the structure of the questionnaire (Klungel et al. 2000).

This thesis covers some of the problems concerning missing data and data collection in psychiatric research; specifically the effects of non-participation, the heterogeneity of study subjects, and data collection methods. The data originate
from a large population-based birth cohort. The aim was to explore how participation in an epidemiologic study that includes questionnaires and a clinical examination is affected by mental health. The purpose was also to study whether non-participants experience more severe clinical symptoms than participants in a psychiatric field study among subjects with a psychosis. Furthermore, other associative factors expected to affect non-participation among subjects with a psychosis were explored. An example of how to improve estimates of brain volumes among subjects with schizophrenia by using inverse probability weighting was presented. Additionally, the precision between self-reported medication use and pharmacy data was explored to assess whether self-reported data are reliable enough for research purposes and whether the prevalence of medication use can be estimated precisely using data collected by postal questionnaire.
2 Review of the literature

2.1 Missing data

2.1.1 Sampling

A sample is a portion or subset of a larger group called a population (Fink 2003). Samples are often used instead of populations in medical research because populations may be too large to study in their entirety. A good sample is representative of the population, i.e. important characteristics are distributed similarly. The value of samples lies in the accuracy with which they represent the target population. The target population consists of subjects or units to whom the findings of the research are to be applied.

Inclusion criteria of a study consist of the characteristics of individuals that make them eligible for participation; exclusion criteria are the characteristics that rule out certain people (Fink 2003). Some subjects are excluded from the analyses because they do not fulfil the inclusion criteria, i.e. they are missing by definition (Acock 2005). Missing values of such subjects are ignorable. If a subject is eligible, i.e. the subject is intended to be sampled, missingness is non-ignorable, i.e. should be considered.

The inclusion and exclusion criteria of subjects should be carefully and completely defined before the sample is collected. Selection of subjects in psychiatric epidemiology may be e.g. age specific (adolescents, adults or elderly) or based on the disorder. For example, when studying schizophrenia, those without certain criteria for schizophrenia, e.g. based on the Diagnostic and Statistical Manual of Mental Disorders (DSM), are excluded. Subjects may also be excluded due to several illness-related factors like use of medication, substance abuse or prior psychiatric illnesses. Studies that involve conducting magnetic resonance imaging (MRI) have some specific exclusion criteria, such as having a cardiac pacemaker or metal implants, or being pregnant.

In sampling, many individuals identified as part of the sample may not participate. This may cause non-response bias. Such bias may occur if non-participants differ systematically from participants. As the response rate decreases, the potential for a biased sample increases.
2.1.2 Types and patterns of missing values

Unit and item non-response

Any data missing from the complete data set, i.e. the total sample intended to be studied, is called missingness. Missingness can be caused by either unit or item non-response (Rubin 1976, Little & Rubin 1987, Schafer & Graham 2002). Unit non-response, or non-participation, occurs when the entire data collection fails, i.e. a subject’s follow-up data are completely missing. It may be that the researchers are not able to contact the subject, the subject may refuse to participate in the research, or the subject may be deceased (Young et al. 2006, Eaton et al. 2007). The first two reasons denote eligible subjects and the last one, ineligible subjects. If a subject participates but does not respond to certain individual items, i.e. partial data of the subject are available, missingness is called item non-response.

Wave non-response

Study subjects in longitudinal studies may be present for some waves of data collection and missing for others, which is called wave non-response (Schafer & Graham 2002). A subject may be missing during one wave but may return to subsequent waves. A special case of wave non-response is attrition, when a subject does not return to participate in the study. Attrition may be divided into loss to follow-up, i.e. the subject cannot be traced to be invited to participate, and dropping out, i.e. the subject no longer wishes to participate in the data collection. In longitudinal research with annual follow-up questionnaires, response rates decline substantially at first, followed by a lesser decrease thereafter (Goldberg et al. 2006).

Types of missingness

Handling of missing values depends on the type of missingness, also called missing data mechanisms. Observations may be missing based on the following mechanisms: missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR) (Little & Rubin 1987). MCAR and MAR are considered ignorable. MCAR does not lead to bias, and auxiliary information may
help explain the potential bias in MAR by adjusting for the covariates related to missingness. MNAR is, on the other hand, considered non-ignorable.

Observations are **MCAR** when the probability that a subject remains in the study does not depend on exposure, confounders, or the outcome (Little & Rubin 1987). Subjects remaining in the study are hence assumed to be a random sample of all eligible subjects. If, for example, one were to study the effect of birth complication on the risk for schizophrenia, adjusted for gender, missingness caused by the MCAR mechanism would not depend on any of these three variables. This mechanism rarely applies in practice, however (Salim et al. 2008).

Secondly, observations are **MAR** when the probability that a subject remains in the study depends on exposure or confounders but not on the outcome (Little & Rubin 1987). In the example above, missingness caused by the MAR mechanism would be related to either birth complication or gender, but not to the risk for schizophrenia.

Finally, observations are **MNAR** when the probability of dropping out depends on the outcome and cannot be completely explained by the covariates (Schafer & Graham 2002, Twisk & de Vente 2002). Missingness caused by MNAR in the example above would suggest that dropping out is dependent on the outcome, i.e. the risk for schizophrenia.

**Patterns of missingness**

There are three different patterns of missingness (Schafer & Graham 2002). In a **univariate pattern** of missingness, certain items are completely observed, but missing values occur in some other item or set of items. In a **monotone pattern**, an item, say $Y_j$, is either observed or not. If item $Y_j$ is missing for a unit, then items $Y_{j+1}$, $Y_{j+2}$, …, $Y_p$ for that unit are missing as well. Finally, in an **arbitrary pattern**, any set of variables may be missing for any unit.
2.1.3 Factors associated with non-participation

Participation in epidemiological studies has been associated with socioeconomic, demographic and health-related factors, and with participation in earlier studies (Table 1). Attrition with respect to psychiatric disorders has been studied in, e.g. the following epidemiological studies: Epidemiologic Catchment Area study (ECA) (Eaton et al. 1992, Badawi et al. 1999, Eaton et al. 2007), Netherlands Mental Health Survey and Incidence Study (NEMESIS) (Bijl et al. 1998, de Graaf et al. 2000), general population study of mental disorders in Stockholm, Sweden (PART Study) (Lundberg et al. 2005), Prenatal Determinants of Schizophrenia Study (Susser et al. 2000), and Finnish longitudinal survey on depression (Eerola et al. 2005).

Socio-demographic factors

In general, men tend to participate in studies less actively than women (Eerola et al. 2005, Lundberg et al. 2005, Carlsson et al. 2006, Harald et al. 2007, Koskinen et al. 2008). In the ECA study male sex was associated with loss of contact but not with refusal at the one-year follow-up (Eaton et al. 1992).

Age has different associations with attrition due to failure to locate rather than due to refusal. In the ECA study, younger age groups were associated with loss of contact at the one-year and 15-year follow-ups (Eaton et al. 1992, Badawi et al. 1999). Older age groups (45–64 years) were associated with refusal at the one-year follow-up (Eaton et al. 1992), but subjects in the oldest age group (over 65 years) were least likely to refuse participation at the 15-year follow-up (Badawi et al. 1999). In the NEMESIS study, risk for attrition due to loss of contact decreased with increasing age, but increased due to refusal in the youngest (18–24 years) and oldest (55–64 years) age groups (de Graaf et al. 2000). In a Study of the Prevalence of Mental Disorder in Oslo, Norway, in a random sample of subjects aged 18–65 years, increasing age was associated with attrition due to both loss of contact and refusal (Kringlen et al. 2001).

Married or cohabiting people usually participate more actively than the unmarried, e.g. in the PART Study (Lundberg et al. 2005), in the 1936 Cohort of Copenhagen, Denmark at the 60-year follow-up (Drivsholm et al. 2006), and in the Health 2000 Survey (Koskinen et al. 2008).
Participants tend to have a higher educational level (Eerola et al. 2005, Lundberg et al. 2005, Carlsson et al. 2006, Drivsholm et al. 2006) and a higher socioeconomic status (Kringlen et al. 2001, Lundberg et al. 2005, Drivsholm et al. 2006, Atherton et al. 2008) than non-participants.

Subjects living in some geographic areas, specifically in urban areas (de Graaf et al. 2000, Kringlen et al. 2001, Koskinen et al. 2008) or areas with a high level of deprivation (Williams et al. 2007), tend to participate less actively. In the Health 2000 Survey, in which the data were collected by personal interviews and health examinations, the participation rate was highest among people living in the catchment area of Oulu University Hospital, and lowest in the Helsinki and Turku areas (Koskinen et al. 2008). There are also differences in participation activity between different countries. Response rates in the European multicentre study of vertebral osteoporosis have been higher in the Nordic countries and lower in Southern Europe (O’Neill et al. 1995). Studies by Lundberg et al. (2005) and Carlsson et al. (2006) have shown that subjects born in the Nordic countries participate more actively than those who have immigrated.

One consistent finding is that earlier participation predicts later participation in longitudinal studies (Drivsholm et al. 2006). Finally, there are also studies suggesting that non-participants may be more similar to participants when the response rate is low (Stang & Jöckel 2004, Carlsson et al. 2006). More specifically, Stang & Jöckel (2004) concluded that if exposure misclassification is substantial and increases with the difficulty of recruiting a study subject, studies with low response rates may in fact be less biased than studies with higher response rates. Carlsson et al. (2006) suggested that if the response rate is high, the non-participant minority might be clearly distinct from the participants and hence, as participation decreases, non-participants may be more similar to participants.

**Health-related factors**

According to Drivsholm et al. (2006), subjects who participate in health studies have better health profiles than those who do not participate. Also Koskinen et al. (2008) stated in the methodology report of the Health 2000 Survey that participation in health surveys is significantly lower among people with severe mental health problems. Population-based studies are the focus of this thesis, but participation with respect to psychiatric illness has been studied in some previous studies on psychiatric patients; e.g. Fischer et al. (2001) studied previously
hospitalised psychiatric patients and Vanable et al. (2002) studied psychiatric outpatients. Riedel et al. (2005) studied the representativeness of schizophrenia inpatients participating in clinical trials. These studies suggest that more psychiatric problems occur among non-participants than among participants.

In the study by Fischer et al. (2001), previously hospitalised psychiatric patients who could not be contacted after several follow-up attempts had been generally more severely impaired during the baseline hospitalisation. Participants had had better prognoses as shown by clinical assessments done at both admission and discharge. In the study by Riedel et al. (2005), psychiatric patients participating in clinical trials were shown to have a shorter duration of illness and a lower number of psychiatric hospitalisations. Also, in the population-based study by Drivsholm et al. (2006), non-participants were shown to have had episodes of psychiatric hospitalisations more often than participants. The number of disorders was strongly associated with both refusal and loss of contact in the one-year follow-up of the ECA study (Eaton et al. 1997). Those with four or more psychiatric disorders were more likely to refuse and less likely to be contacted. In the study by Riedel et al. (2005), participants more likely had atypical, and less likely had typical or a combination of atypical and typical antipsychotic medication on discharge, compared with non-participants.

De Graaf et al. (2000) stated that psychopathology has a limited effect on overall attrition. Bijl et al. (1998) also found that non-participants did not differ significantly from participants in estimated psychiatric morbidity. However, they suggested that people suffering from schizophrenia may be less willing or able to participate in interviews. In the NEMESIS study, people with no fixed address, insufficient proficiency in Dutch or long periods of institutionalisation were not reached. According to the authors, this could have influenced the effect of e.g. schizophrenia on participation. On the other hand, in studies by Susser et al. (2000) and Lundberg et al. (2005), schizophrenia was associated with non-participation.

Non-participants have also been shown to have schizotypal or delusional disorders in a population-based study (Lundberg et al. 2005) and psychoactive or adjustment disorders in an outpatient study (Vanable et al. 2002) more often than participants. Despite schizotypal disorders being associated with non-participation according to Lundberg et al. (2005), Susser et al. (2000) suggested schizotypal patients are more likely to participate. Within schizophrenia, patients with residual schizophrenia may participate more and patients with disorganized
schizophrenia less according to Cheung et al. (1997), who studied attrition bias in a study of aggressive behaviour in schizophrenia patients.

Attrition is not related only to schizophrenia. Participants have been shown to more likely have bipolar disorder (Fischer et al. 2001), affective disorder (Susser et al. 2000), or anxiety or mood disorders (Fischer et al. 2001) than non-participants. In a study by Vanable et al. (2002), non-completers were more likely to have an adjustment disorder than those who completed the study. According to Lundberg et al. (2005), non-participants more likely have substance use disorders than participants.

In terms of other health-related factors, those who participate in studies have been shown to have better self-rated health, among men (Thomas et al. 2002) and among women (Young et al. 2006), and to be less often on disability pension (Lundberg et al. 2005). It has also been noted that non-participants have more unhealthy behaviour, e.g. smoking (Drivsholm et al. 2006, Atherton et al. 2008).
Table 1. Variables associated with participation in previous population-based studies.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Label</th>
<th>Association with participation and type of attrition when specified</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous participation</td>
<td>Participation in previous phases of longitudinal study</td>
<td>Higher participation</td>
<td>Drivsholm et al. 2006</td>
</tr>
<tr>
<td>Socio-demographic factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Younger</td>
<td>Higher rates of loss of contact</td>
<td>Eaton et al. 1992</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Higher rates of refusal</td>
<td>de Graaf et al. 2000</td>
</tr>
<tr>
<td></td>
<td>Older</td>
<td>Higher rates of loss of contact</td>
<td>Kringlen et al. 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lower rates of loss of contact</td>
<td>de Graaf et al. 2000</td>
</tr>
<tr>
<td>Living area</td>
<td>Urban</td>
<td>Lower participation</td>
<td>de Graaf et al. 2000, Kringlen et al. 2001</td>
</tr>
<tr>
<td></td>
<td>High deprivation level</td>
<td>Lower participation</td>
<td>Williams et al. 2007</td>
</tr>
<tr>
<td></td>
<td>Middle-class suburbs</td>
<td>Higher participation</td>
<td>Kringlen et al. 2001</td>
</tr>
<tr>
<td>Variable</td>
<td>Label</td>
<td>Association with participation and type of attrition when specified</td>
<td>References</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------------</td>
<td>---------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Health-related factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychiatric morbidity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>More psychiatric problems</td>
<td>Lower participation</td>
<td>Fischer et al. 2001, Vanable et al. 2002</td>
</tr>
<tr>
<td></td>
<td>More disorders</td>
<td>Higher rates of attrition</td>
<td>Eaton et al. 1992</td>
</tr>
<tr>
<td>Severity of illness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shorter duration of illness</td>
<td>Higher participation</td>
<td>Riedel et al. 2005</td>
</tr>
<tr>
<td></td>
<td>Lower number of psychiatric hospitalisations</td>
<td>Higher participation</td>
<td>Riedel et al. 2005, Drivsholm et al. 2006</td>
</tr>
<tr>
<td>Medication use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Atypical</td>
<td>Higher participation</td>
<td>Riedel et al. 2005</td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schizophrenia</td>
<td>Lower participation</td>
<td>Susser et al. 2000, Lundberg et al. 2005</td>
</tr>
<tr>
<td></td>
<td>Residual SZ</td>
<td>Higher participation</td>
<td>Cheung et al. 1997</td>
</tr>
<tr>
<td></td>
<td>Disorganised SZ</td>
<td>Lower participation</td>
<td>Cheung et al. 1997</td>
</tr>
<tr>
<td></td>
<td>Negative symptoms</td>
<td>Lower participation</td>
<td>Riedel et al. 2005</td>
</tr>
<tr>
<td>Disorder</td>
<td>Schizotypal or delusional DO</td>
<td>Lower participation</td>
<td>Lundberg et al. 2005</td>
</tr>
<tr>
<td></td>
<td>Schizotypal DO</td>
<td>Higher participation</td>
<td>Susser et al. 2000</td>
</tr>
<tr>
<td></td>
<td>Bipolar DO</td>
<td>Higher participation</td>
<td>Fischer et al. 2001</td>
</tr>
<tr>
<td></td>
<td>Affective DO</td>
<td>Higher participation</td>
<td>Susser et al. 2000</td>
</tr>
<tr>
<td></td>
<td>Anxiety or mood DO</td>
<td>Higher participation</td>
<td>Fischer et al. 2001</td>
</tr>
<tr>
<td></td>
<td>Psychoactive or adjustment DO</td>
<td>Lower participation</td>
<td>Vanable et al. 2002</td>
</tr>
<tr>
<td></td>
<td>Substance use DO</td>
<td>Lower participation</td>
<td>Lundberg et al. 2005</td>
</tr>
<tr>
<td>Suicidility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>More suicidal</td>
<td>Lower participation</td>
<td>Vanable et al. 2002</td>
</tr>
<tr>
<td>Prognosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Better prognosis</td>
<td>Higher participation</td>
<td>Fischer et al. 2001</td>
</tr>
<tr>
<td>Disability pension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Being on a disability pension</td>
<td>Lower participation</td>
<td>Lundberg et al. 2005</td>
</tr>
<tr>
<td>Unhealthy behaviours</td>
<td>Smoking</td>
<td>Lower participation</td>
<td>Drivsholm et al. 2006, Atherton et al. 2008</td>
</tr>
</tbody>
</table>

DO = Disorder, SZ = Schizophrenia. ^ Patient study.
2.1.4 Retaining response rates in studies

According to Morton et al. (2006), the response rates in epidemiologic cohort studies have declined in recent years. Altman (2000) proposed an 80 percent response rate as a minimum in acceptable research. Kristman et al. (2004) found considerable bias with a response rate as high as 80% when the missing data mechanism was MNAR. When the missing data mechanism was MAR, they found an unbiased estimate of effect with a response rate of 40%.

Scott et al. (2006) suggest that the first step in retaining participants in longitudinal studies is to identify the reasons why subjects who have participated in at least one wave of data collection do not participate in follow-up data collections. Illness, injury, or institutionalisation may cause inability to participate. Refusal to participate may be due to a lack of trust, insufficient engagement with the research team, inconvenient research appointments, or unpleasant experiences during previous waves of data collection. Finally, contact problems may be due to a change of address and a lack of alternative contact information on how to contact the research subject.

Scott’s model for retaining participants in longitudinal studies (Scott et al. 2006) encompasses the following components: delineation of the roles and responsibilities of the research team; engagement of institutions that interact with research participants; development and use of appropriate materials for education, consent, and tracking participants; development and implementation of a protocol for maintaining all the subjects in the study; monitoring of the research team’s compliance with the preceding protocol; and finally, facilitation of regular case review meetings by the research team.

In the review article by Edwards et al. (2002), the authors identified strategies for increasing the response rates of postal questionnaires. They found that contacting the study subjects before they received the questionnaire and following-up the study subjects increased response rates, as did asking for reasons for not participating. Both monetary and non-monetary incentives increased response rates. Shorter length and e.g. interesting content of questionnaires were also shown to increase response rates. The importance of informed consent should also be noted. When the study, with the risks and benefits of participation, is clearly explained to the subject, participation is less likely to suffer from uncertainty issues (Williams et al. 2007).
2.1.5 Reporting of missing data

Harris et al. (2009) reported common statistical and research design problems in manuscripts submitted to high-impact psychiatry journals. One of the requirements by editors and reviewers was that the authors should clearly describe the extent of missing data. Missing data should be treated properly in the analyses by using statistically appropriate methods. In the review article by Klebanoff & Cole (2008), the use of advanced methods to handle the missing data problem in epidemiologic literature was evaluated. They found that in less than two percent of the articles they examined, the authors had used multiple imputation, inverse probability weighting, or an expectation-maximisation algorithm. They noted, though, that perhaps the listed methods are used more commonly than their survey found, but their use is still quite rare.

Similar conclusions were made in a study by Albon et al. (2008), who evaluated structural neuroimaging studies on psychosis in their systematic review. They reviewed 24 studies that evaluated the clinical benefit of computed tomography, structural MRI or their combinations in psychotic patients. In studies using computed tomography, descriptions of study population selection criteria were considered generally poor, and in MRI studies, adequate. The authors pointed out that in many studies the authors did not give any reason for patients not having been scanned. Also, it was not clear in all the studies how recruitment of the sample had been conducted. Albon et al. (2008) stated that sampling bias is likely to affect the results of all the included studies, and their internal validity is questionable.

Reporting of non-participation has also been generally poor in many of the cohort or population-based studies with MRI scans of the brain among psychotic subjects published recently, during 2008–2010. Most of them did not report patient selection or the extent of non-participation at all. Reporting of non-participation was an exception. For example, Price et al. (2010) stated that about half of the eligible subjects were scanned and compared those with scans to those without. Many of the studies had references to previous articles, which, however, did not necessarily have much information on non-participants. On the other hand, Brown et al. (2009) had references to Susser et al. (2000), who thoroughly evaluated non-participation in the Prenatal Determinants of Schizophrenia study.
2.2 Statistical methods in handling missing data

Missing data reduce sample size and statistical power, and may introduce non-response bias. Differences in outcome variables between cases and controls may be affected by non-participation, as those with certain risk factors may cumulate into participants or non-participants. Non-response bias may occur especially if the missing data mechanism is MNAR.

Most notably, statisticians Donald Rubin, Roderic Little and Joseph Schafer have developed a classification scheme for missing data and methods to deal with it. According to Little & Rubin (1987), the literature on the analysis of partially missing data is comparatively recent, and has flourished since the early 1970s. The following chapters are mainly based on Rubin (1987), Little & Rubin (1987), Schafer & Graham (2002), and Haukoos & Newgard (2007).

2.2.1 Procedures based on completely recorded units

A widely used method known as complete-case analysis (or listwise deletion) limits the analysis to observations for which all values are observed. This method is used by default in many statistical programs, but it has substantial and important disadvantages. Due to the reduction in the sample size, the precision of reported estimates decreases. Complete-case analysis is also likely to bias the estimates, unless the missingness is MCAR. Complete-case analysis should be used only if the proportion of complete cases is high and the potential for bias and loss of precision can be believed to be minimal. (Little & Rubin 1987, Schafer & Graham 2002, Haukoos & Newgard 2007.)

A form of complete-case analysis is available-case analysis (or pairwise deletion), which uses sets of units for different parameters, i.e. it deletes only observations with missing variable values. According to Schafer & Graham (2002), the underlying principle of making use of all available data is sensible, but deletion of missing values is a poor way to operationalise it. Available-case analysis is used in univariate or bivariate statistics in which complete-case analysis would delete subjects with some available data. The disadvantages of available-case analysis include variation in the number of subjects used in different analyses, a reduction in the precision of estimates based on different sets of variables compared, and the potential for bias. (Little & Rubin 1987, Schafer & Graham 2002, Haukoos & Newgard 2007.)
Averaging the available items (or case-by-case deletion) of the individual is used in situations where characteristics of interest are measured by scales rather than by single items. In psychiatry, various different rating scales are typically used to measure outcomes. If one or more items are missing from the scale, averaging the available items seems more reasonable than deleting all the items of a subject. This method is largely unstudied. Even though averaging the available items may theoretically introduce bias even under MCAR, in practice this method generally leads to unbiased results. (Schafer & Graham 2002.)

### 2.2.2 Single imputation

Imputation is a generic term for filling in missing data with plausible values (Schafer 1997). In imputation methods, missing items are filled in with their approximates. Analysis does not delete any subjects, thus it prevents loss of power caused by an otherwise diminished sample size (Schafer & Graham 2002, Haukoos & Newgard 2007). Imputation methods can also retain high precision if the observed data contain adequate information for predicting the missing values. Finally, if the data are to be analysed by multiple researchers, imputing prior to any analyses ensures that the same data are used in all the analyses. According to Schafer & Graham (2002), single imputation is reasonable, and superior to complete-case analysis, if the response rate is high but the missing pattern is arbitrary. The negative sides of single imputation are noted in the methods presented below as they have been stated by Haukoos & Newgard (2007).

The most basic forms of single imputation are mean or median imputations, which replace missing values with means or medians of the available data. This method leads to biased estimates, because missing values are replaced by values from the centre of the distribution, decreasing the variance of the estimates.

In regression imputation missing values are replaced by single predicted values from the regression analysis using available data. This method may generate reasonable approximations for the missing values, but it also underestimates variance. Stochastic regression imputation adds a residual error to the approximation from regression analysis, which is imputed into the missing values. The quality of both regression imputation methods depends heavily on whether the regression model is correct, i.e. the selection of predictors is appropriate.

In hot deck imputation a missing value is replaced by the value of a subject most similar to the one with the missing value. This method also underestimates
the variance of the parameter estimate, but it does not require parametric assumptions or careful modelling to identify the values to impute. Hot deck imputation needs a large enough sample to find an adequate subject for the imputed value. Cold deck imputation works like hot deck imputation, but the imputed value is collected from an external data source. This method has not been widely adopted and is generally not recommended, because external data may differ systematically from the primary data.

In longitudinal studies, the last observation carried forward method uses the last observed value of the item, assuming that the value did not change from the previous measurement. This method is likely acceptable if measurements are expected to be relatively constant over time. In worst (or best) case analysis the worst (or best) case, e.g. death (or recovery), is imputed for the missing values in case of e.g. binary survival outcome, to define the range of possible results under a worst (or best) case scenario.

2.2.3 Multiple imputation

In multiple imputation (MI) each missing value is replaced by two or more acceptable values representing a distribution of possibilities; this idea was originally proposed by Rubin (1977, 1978). Observed values are used to generate a range of plausible values for each missing value, based on existing correlations between variables (Rubin 1987).

MI is an extension of single imputation. Missing values are replaced by \( m > 1 \) simulated values, where \( m \) is typically small (say, 3–10), creating \( m \) complete data sets (Rubin 1987, Schafer 1999). The simulated complete data sets are analysed using standard methods, and the results are combined in order to produce estimates and confidence intervals that incorporate missing data uncertainty. Multiply imputed data are intended to draw inferences from (e.g. point estimates and confidence intervals) rather than assess any individual data.

The MI method is simple in terms of analyses. Once the simulations are done the data may be analysed by virtually any technique appropriate for complete data. According to Newgard & Haukoos (2007), MI is computationally straightforward, versatile, relatively easy to apply, and increasingly available in standard statistical software applications. However, the method depends heavily on the generation of imputations (Schafer 1999). Also Philipson (2008) noted that MI is a likelihood-based approach, which requires correct specification of the imputation model.
Imputations should, on average, give reasonable predictions for the missing data, and the variability among them should reflect an appropriate degree of uncertainty.

2.2.4 Weighting methods

Weighting can be used to correct non-response bias due to unit non-response. It is the most appropriate approach for scenarios where entire data sets are missing, except for e.g. some basic socio-demographic information (Little & Rubin 1987, Höfler et al. 2005). These auxiliary variables are not of interest in the analyses themselves, but rather background information on all eligible subjects. Auxiliary variables can be e.g. gender, age, or prior medical condition collected at a baseline investigation or from register data.

In weighting methods in general, complete cases are weighted so that they more closely resemble the full sample with respect to auxiliary variables. Adjustment of weights to account for non-participation is done using information from the auxiliary variables of all eligible subjects. Corresponding data on the auxiliary variables for both participants and non-participants are therefore needed. Weighting can correct for bias related to the auxiliary variables used to model the response probabilities, but not to unused or unmeasured variables. Therefore, the quality of weighting depends on the quality of the auxiliary variables and the model built on them. (Schafer & Graham 2002.)

Selection probability \( P_s \) is calculated as the ratio between the number of subjects selected in the study \( n \) and the number of all eligible subjects in the population \( N \), i.e.

\[
P_s = \frac{n}{N}.
\]

The so-called base weight \( w_B \) for an individual \( i \) (\( i = 1, \ldots, N \)) is the inverse of his or her selection probability, i.e.

\[
w_B = \frac{1}{P_s}.
\]

Weight adjustments are employed to adjust base weights so that the weighted distributions for study variables in the sample conform to the target population distribution for these variables. (Kalton & Flores-Cervantes 2003.)
Weighting class method

The weighting class method (or cell weighting) can be used if the missing data mechanism is MAR and the auxiliary variables are categorical. In this method, weighting classes that are homogeneous with respect to auxiliary variables and propensity for responding are created (Little & Rubin 1987). In each weighting class, i.e. cell $j$ ($j = 1, \ldots, J$), a weight is computed for all participants by the inverse of the response probability ($P_{R_i}$) of subject $i$ multiplied by the average response probability ($\bar{P}_{R_i}$) within all participating subjects in cell $j$, i.e.:

$$w_{i,j} = \frac{1}{P_{R_{i,j}}} \times \bar{P}_{R_i},$$

where $w_{i,j}$ is the weight for participating subject $i$ in cell $j$.

One disadvantage of the weighting class method is that it can lead to large variability in the distribution of weighting adjustments. It is of particular concern when there are many weighting classes with small sample sizes. (Kalton & Flores-Cervantes 2003.) In order to create stable adjustments, the data need a reasonable (at least 20) number of participants per weighting class, and fewer non-participants than participants per weighting class (Carlson & Williams 2001, Potter et al. 2006).

Post-stratification is similar to the weighting class method except that population counts are used to adjust the weights (Little 1993). The method of raking uses estimates of cell counts that satisfy marginal constraints (Little & Rubin 1987).

Inverse probability weighting

The use of propensity scores (Rosenbaum & Rubin 1983, Rosenbaum & Rubin 1984) is more practical when there are a large number of auxiliary variables to consider. Cassel et al. (1983) defined weights as the inverses of the estimated propensity scores. In this method, namely inverse probability weighting (IPW), propensity to participate is estimated using all available and relevant data from the auxiliary variables.

Let $R_i$ be the indicator for whether or not unit $i$ would respond if selected into the sample:
Response propensity can be estimated through a logistic regression model, i.e.

\[ \Pr(R = 1 \mid X) = \frac{1}{1 + \exp[-(a + bX)]}, \]

where \( \Pr(R = 1 \mid X) \) denotes the probability of a response, \( X \) the auxiliary variables, and \( a \) and \( b \) the unknown regression coefficients. This model is applied to participating subjects, and a probability of responding is generated for each subject. The adjusting weight \( w_i \) for unit \( i \) is then computed analogously with the weighting class method, i.e. by

\[ w_i = \frac{\Pr(R = 1 \mid X)}{\Pr(R = 1 \mid X)_i}, \]

where \( \Pr(R = 1 \mid X)_i \) is the average response propensity within all participating subjects.

According to Little & Rubin (1987), IPW removes non-response bias, but it may yield estimates with high variance. Participants with very low estimated response propensity receive large weights and may be unduly influential in the estimation of outcome variables. Also, weighting by estimated propensity scores depends highly on the correct model specification.

### 2.3 Precision between self-reported and pharmacy data on medication use

#### 2.3.1 Self-reported data on medication use

The generalisability of research results is affected not only by non-participation but also the quality of the collected data. Self-report measures on medication use are comprised of questionnaires, diaries, and interviews. Many population studies are based on data collected by postal questionnaire, because it is a reasonably simple and inexpensive way of collecting data from large samples (Edwards et al. 2002). Postal questionnaires can be used to collect data e.g. on medication use and health. In the literature review on medication adherence by Garber et al. (2004), self-report questionnaires and diaries exhibited moderate to high
concordance with e.g. administrative claims or clinical opinion, whereas personal interviews did not perform as well.

To the knowledge of the author, publications on the reliability of self-reported measures on psychoactive medication use are scarce. One of the first studies evaluating the validity of register-based information on psychoactive drug utilisation has been conducted by Haukka et al. (2007), who compared interview data with prescription reimbursement data in a population-based study of schizophrenia. Also other studies comparing self-reported data on medication use and pharmacy records in large, population-based studies are scarce (Monster et al. 2002, Caskie et al. 2006, Nielsen et al. 2008). In many studies the accuracy of self-reported medication use has been studied in selected populations, in e.g. elderly (Guénette et al. 2005), adolescents (Skurtveit et al. 2008), or inpatients (Glintborg et al. 2007).

The reliability of data obtained with self-report inquiries depends on the ability and will of the subjects to recall information accurately (West et al. 1995, Boudreau et al. 2004, Caskie & Willis 2004), as well as on the data collection methods and the structure of the questionnaire (Klungel et al. 2000, Lunet et al. 2008). According to Carlsson et al. (2006), self-reported questionnaires may have less validity than interviews and medical examinations. In the study by Klungel et al. (1999), the accuracy of data was better for questions about drugs used for a specific indication than for open-ended questions.

Recall of medication use varies with the type of drug. Monster et al. (2002) and Caskie et al. (2006) have shown that agreement between questionnaire data and pharmacy records is good for chronically used drugs. More specifically, Monster et al. (2002) found that agreement between questionnaire data and national prescription records in a population-based Dutch cohort was good for chronically used drugs but not very good for drugs used for shorter periods, drugs used as needed, or drugs that were available over the counter. In the population-based Seattle Longitudinal Study, Caskie et al. (2006) found that information recorded using the brown bag data collection method for drugs used for serious conditions or on a regular basis is congruent with national prescription records.

The agreement between personal interview and register-based data has been shown by Haukka et al. (2007) to be good for most psychotropic drugs. In their study agreement was best for lithium use, followed by antipsychotic drug use and antidepressant use. Nielsen et al. (2008) found substantial agreement for antipsychotic and antidepressant medications in their study comparing interview data with national prescription records in a large, population-based Danish health
survey. In a study of adolescents, Skurtveit et al. (2008) found better specificity for psychoactive medications compared with some other medications, but in their study psychotropic drugs were dispensed to only eight people. Kwon et al. (2003) found good agreement between self-reported questionnaire data and pharmacy data on antidepressant use in a longitudinal depression study.

Gender has not been shown to influence recall of drug consumption, whereas increasing age decreases the level of recall (van den Brandt et al. 1991, West et al. 1995). For example, according to Guénette et al. (2005), interviews of elderly people exhibited poor agreement with pharmacy records. Having high household income, being married and having good health have been associated with better self-reporting (Cotterchio et al. 1999, Caskie & Willis 2004). In the study by Klungel et al. (2000), those with lower education recalled medication use worse than those with higher education.

2.3.2 Register data on medication use

The Finnish register system provides good opportunities for utilising data in epidemiological research (Gissler & Haukka 2004). A personal identification code has been allocated to all Finnish citizens since the 1960s. It enables linkage of data between different registers and also between registers and other data under special legislation. Chapter 4, Section 14, of the Personal Data Act (Personal Data Act 1999) concerning processing of personal data for special purposes states that personal data may be processed for historical or scientific research purposes, if:

- the research cannot be carried out without data that identifies persons and the consent of the register subjects cannot be obtained owing to the quantity of the data, the age of the data or some other comparable reason;
- the use of the personal data file is based on an appropriate research plan and the person or group of persons responsible for the research have been designated;
- the personal data file is used and the personal data therein are disclosed only for historical or scientific research purposes and the procedure followed is also otherwise such that data pertaining to a given individual are not disclosed to outsiders; and
- after the personal data are no longer required for the research or for verification of the results achieved, the personal data file is destroyed or
transferred into an archive, or the data therein are altered so that the register subjects can no longer be identified.

According to Miettunen et al. (in press), register studies have strengths in terms of statistical power and representativeness, as well as enabling examination of issues otherwise difficult to study, such as rare diseases. Finnish, as well as other Nordic registers, are based on the whole population, and they have good coverage and validity (Gissler & Haukka 2004, Miettunen et al. in press).

The Social Insurance Institution of Finland manages National Health Insurance in Finland (Social Insurance Institution of Finland 2008). All official residents of Finland are entitled to medication reimbursement. Reimbursements for pharmaceuticals are based on the severity and duration of illness. Medications for severe conditions such as cancer, type 1 diabetes, and psychotic and bipolar I disorders are fully reimbursed. Medications for some other chronic conditions such as hypertension are 75% reimbursed. The Finnish National Prescription Register covers all pharmacies in Finland, and it registers all drug purchases for which the Social Insurance Institution of Finland has paid any reimbursement.

Pharmacy data constitute an easily obtainable tool in epidemiological studies (Monster et al. 2002, Gissler & Haukka 2004). Miettunen et al. (in press) note that prescription registers and other sources of administrative data have become an important source of information in pharmacoepidemiological studies. Data from prescription registers can be used to study medication patterns in large populations.

Methods for capturing medicine users are usually based on either daily defined dose (DDD) or assigning a fixed period of usage to each prescription record. Legend duration time is calculated by dividing the number of units dispensed by the DDD (Mantel-Teeuwisse et al. 2001). In the legend duration time method, legend duration time is used to define a subject as a prevalent medication user if the legend duration time from the date of drug purchase exceeds the date of estimated point prevalence. In the fixed-time method, subjects are assumed to be prevalent medication users if they have purchased drugs within a certain specific period of time preceding the date of estimated point prevalence. The legend duration time method outlines the medicines for which DDDs are not assigned (Nielsen et al. 2008). In the study by Nielsen et al. (2008), the 90-day fixed-time method was shown to work better than the legend time method in capturing the use of in-need medications, but the two methods did not differ in capturing the use of medications for chronic disorders.
2.4 Agreement statistics between two data sets

Agreement statistics are usually used to test reliability between diagnostic tests (e.g. Viera & Garrett 2005). In survey analysis, reliability is a measure of reproducibility of the survey instrument or test. In this chapter agreement statistics are presented as means of comparing self-reported medication use and pharmacy data, as in original study IV.

Agreement statistics can be calculated from a fourfold table, as shown in Table 2. Let ‘+’ indicate the presence of medication use and ‘-’ indicate the absence of medication use in the two data sets. The letters a-d denote the actual numbers of units in each cell. The marginal totals are \( f_1, f_2, g_1 \) and \( g_2 \).

<table>
<thead>
<tr>
<th>Pharmacy data</th>
<th>+</th>
<th>-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-reported data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>a</td>
<td>b</td>
<td>( f_1 )</td>
</tr>
<tr>
<td>-</td>
<td>c</td>
<td>d</td>
<td>( f_2 )</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>( g_1 )</td>
<td>( g_2 )</td>
<td>( N )</td>
</tr>
</tbody>
</table>

Table 2. 2×2 table, representing the actual numbers of units in each cell, with marginal totals.

2.4.1 Proportions of agreement

One way of gauging agreement between two data sets is to calculate the proportion of observed agreements \( (p_o) \), which is the proportion of units with the same value in both data sets (Cohen 1960). It is computed by

\[
p_o = \frac{a + d}{N}.
\]

Standard error for \( p_o \) can be calculated by using the standard methods applicable to proportions, i.e.

\[
SE(p_o) = \sqrt{\frac{p_o(1-p_o)}{N}}
\]

and the 95% confidence interval for \( p_o \) by

\[
p_o \pm 1.96 \times SE(p_o),
\]
where 1.96 is the Standard Normal deviate (z) corresponding to the 95% level of
confidence. The other values of z at other, often used levels of confidence are as
follows: z = 1.645 for 90% confidence and z = 2.576 for 99% confidence.

This proportion may be informative, but it has some limitations. For example,
it does not distinguish between agreements on positive or negative ratings.
According to Cohen (1960), the proportion of observed agreement can be high
even with equivalent data rated purely by chance. One can alternatively compute
proportions of specific agreement (Spitzer & Fleiss 1974, Fleiss 1981), which are
agreements relative to each rating category individually.

Proportion of positive agreement ($p_{pos}$) is computed by

$$p_{pos} = \frac{2a}{2a + b + c} = \frac{2a}{f_1 + g_1}.$$ 

Proportion of negative agreement ($p_{neg}$) is computed equivalently by

$$p_{neg} = \frac{2d}{2d + b + c} = \frac{2d}{f_2 + g_2}.$$ 

Asymptotic (large sample) standard errors for $p_{pos}$ and $p_{neg}$ can be estimated by
(Mackinnon 2000)

$$SE(p_{pos}) = \sqrt{\frac{4a(b+c)(a+b+c)}{(f_1 + g_1)^2}}$$

and

$$SE(p_{neg}) = \sqrt{\frac{4d(b+c)(b+c+d)}{(f_2 + g_2)^2}}.$$ 

Asymptotic 95% confidence intervals for $p_{pos}$ and $p_{neg}$ can be obtained as for $p_o$,
substituting $p_{pos}$ and $p_{neg}$ for $p_o$, and using the asymptotic standard errors given
above.

The proportion of positive agreement estimates conditional probability, given
that in one randomly selected data set having a positive rating, the other data set
also has a positive rating. The proportion of negative agreement equivalently
estimates conditional probability, given that in one randomly selected data set
having a negative rating, the other data set also has a negative rating. $p_{pos}$ can be
considered analogous to sensitivity, and $p_{neg}$ to specificity (Cicchetti & Feinstein
1990). The values of $p_{pos}$ and $p_{neg}$ range from 0 to 1. Values close to unity signify
good agreement and values close to zero poor agreement. If both $p_{pos}$ and $p_{neg}$ are fairly large, there is less need to compare agreement by chance.

### 2.4.2 Cohen’s kappa

If, when comparing two data sets, one can be considered a gold standard, *accuracy* can be evaluated. However, when comparing two data sets, neither of which is a gold standard, the right term is *precision* (Viera & Garrett 2005). Accuracy is, therefore, agreement with the gold standard, whereas precision means good reliability between the two data sets. Data can be both accurate and precise.

Precision between two data sets can be reported using *Cohen’s kappa* ($\kappa$) (Cohen 1960), which is intended to give a quantitative measure of the magnitude of agreement between the two data sets. $\kappa$ is computed using $p_o$ (as above) and the *proportion of expected agreements* ($p_e$), which is computed by

$$p_e = \frac{f_1 \times g_1 + f_2 \times g_2}{N^2}.$$  

Cohen’s kappa can then be calculated by

$$\hat{\kappa} = \frac{p_o - p_e}{1 - p_e}.$$  

Standard error for Cohen’s kappa can be calculated by

$$SE(\hat{\kappa}) = \sqrt{\frac{p_o(1 - p_o)}{N(1 - p_e)^2}}$$  

and the 95% confidence interval by

$$\hat{\kappa} \pm 1.96 \times SE(\hat{\kappa}).$$  

The possible values of $\kappa$ range from -1 to 1, though usually they fall between 0 and 1. Unity represents perfect agreement and zero means agreement no better than that expected by chance. Values of $\kappa$ below zero indicate agreement worse than that expected by chance. A commonly cited scale used when interpreting $\kappa$ is as follows: $< 0$, poor (less than chance); 0.00–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect agreement (Landis & Koch 1977). According to Sim & Wright (2005), $\kappa$ does not indicate
whether disagreement is due to chance or a consistent pattern, and the data should be examined accordingly.

### 2.4.3 Effect of prevalence and bias

There are factors that can influence interpretation of $\kappa$ (Byrt et al. 1993, Sim & Wright 2005); namely prevalence and bias. The *prevalence index* ($PI$) indicates the effect of prevalence, i.e. the distribution of outcome, whereas the *bias index* ($BI$) measures the effect of bias, i.e. systematic differences between two data sets. The $PI$ and $BI$ are calculated by

$$PI = \frac{a - d}{N}$$

and

$$BI = \frac{b - c}{N}.$$  

The $PI$ and $BI$ range from -1 to 1. The $PI$ equals zero when the observed agreement equals 50%, negative when prevalence is less than 50% and positive when prevalence is greater than 50%. Specifically, the $PI = -1$ when $a = 0$ and $d = N$, and the $PI = 1$ when $a = N$ and $d = 0$. The $BI$ equals zero when the marginal totals are equal, i.e. $f_1 = g_1$ and $f_2 = g_2$. The absolute value of the $BI = 1$ when either $b = N$ or $c = N$. A negative or positive $BI$, therefore, indicates a higher proportion of positive cases in one of the two data sets.

If the $PI$ is high, i.e. the prevalence of a positive rating is either very high or very low, $\kappa$ reduces (Feinstein & Cicchetti 1990). The effect of prevalence on $\kappa$ is greater for larger values of $\kappa$ than for small values (Byrt et al. 1993). According to Sim & Wright (2005), the $PI$ provides an indirect indication of true prevalence in the population.

Even in situations where the two data sets agree on the proportion of cases, the $BI$ can differ due to the two data sets differing on the proportion of positive cases (Sim & Wright 2005). The $BI$ is low if disagreement is close to symmetrical, i.e. the proportions of positive cases are close to each other. On the other hand, if disagreement is asymmetrical, i.e. the proportions of positive cases differ, the $BI$ increases. When there is a large bias, i.e. the $BI$ is high, $\kappa$ is higher than when bias is low or absent. In contrast to prevalence, the effect of bias is greater for smaller values of $\kappa$ than for large values. (Byrt et al. 1993.)
2.4.4 Prevalence-adjusted and bias-adjusted kappa

Because $\kappa$ is affected by both overall prevalence and any bias between the two data sets, the well-known paradoxes related to $\kappa$ may occur (Feinstein & Cicchetti 1990). The paradoxes described by Feinstein & Cicchetti (1990) are:

“The first paradox of $\kappa$ is that if $p_o$ is large, the correction process can convert a relatively high value of $p_o$ into a relatively low value of $\kappa$… Thus, with different values of $p_o$, the $\kappa$ for identical values of $p_o$ can be more than twofold higher in one instance than in the other.”

“The second paradox occurs when unbalanced marginal totals produce higher values of $\kappa$ than more balanced totals.”

A version of $\kappa$ that adjusts for prevalence and bias is called prevalence-adjusted and bias-adjusted kappa (PABAK) (Byrt et al. 1993). In order to compute PABAK, the actual units in discordant cells, i.e. $b$ and $c$, are replaced by their average, say $m$,

$$m = \frac{b + c}{2}$$

and the actual units in concordant cells, i.e. $a$ and $d$, by their average, say $n$,

$$n = \frac{a + d}{2}.$$

Due to PABAK using $m$ and $n$ instead of $a$, $b$, $c$, and $d$, the marginal totals are all equal, i.e.

$$m + n = \frac{b + c}{2} + \frac{a + d}{2} = \frac{N}{2}.$$

From this $p_e$ can be seen to equal 0.5

$$p_e = \frac{N \times \frac{N}{2} + \frac{N}{2} \times \frac{N}{2}}{N^2} = \frac{2 \times N^2}{4} = \frac{1}{2} = 0.5$$
Therefore, $PABA\mathcal{K}$ can be computed as

$$PABA\mathcal{K} = \frac{2n}{N} - p_e = \frac{N}{N} - 0.5 = \frac{p_o}{0.5} - 1 = 2p_o - 1.$$ 

The possible values of $PABA\mathcal{K}$ range from -1 to 1. Specifically, $PABA\mathcal{K} = -1$ when $n = 0$; $PABA\mathcal{K} = 1$ when $m = 0$; and $PABA\mathcal{K} = 0$ when observed agreement is equal to 50%. Byrt et al. (1993) has shown that $\kappa$ is related to $PABA\mathcal{K}$ by the following formula:

$$\kappa = \frac{PABA\mathcal{K} - PI^2 + BI^2}{1 - PI^2 + BI^2}.$$ 

From this formula it can be seen that if $PABA\mathcal{K} = 1$, then $\kappa = 1$. Otherwise, the larger the absolute value of the $PI$, the smaller is $\kappa$; and the larger the absolute value of the $BI$, the larger is $\kappa$. Depending on the relative size of the $PI$ and $BI$, $\kappa$ may be larger or smaller than $PABA\mathcal{K}$.
3  Aims of the study

The purpose of this study, which utilised the Northern Finland 1966 Birth Cohort data, was to study non-response and information bias in psychiatric research. The study aimed to:

1. Explore how participation in an epidemiologic health study that includes questionnaires and a clinical examination is affected by mental health (original study I). Specifically, the purpose was to test the following hypotheses: Subjects with psychiatric disorders are less likely to participate than those without, and the participation rate depends on the type of psychiatric disorder.

2. Test the hypothesis that the type of psychosis is linked to participation in a field study including magnetic resonance imaging of the brain, psychiatric interviews and cognitive testing among subjects with a psychosis, and to explore other associative factors expected to affect participation (original study II).

3. Present an example of using inverse probability weighting to adjust for non-participation in estimates of grey matter, white matter and cerebrospinal fluid volumes among subjects with schizophrenia (original study III).

4. Compare self-reported medication use and pharmacy data on major psychoactive medications and three classes of medications used for non-psychiatric indications in order to assess whether the data are reliable for research purposes and whether the prevalence of medication use can be estimated reliably (original study IV).
4 Material and methods

4.1 Study population: The Northern Finland 1966 Birth Cohort

This study is based on the prospective, longitudinal Northern Finland 1966 Birth Cohort (NFBC 1966). The NFBC 1966 was originally assembled by Professor (Emerita) Paula Rantakallio, whose purpose was to examine the risk factors for perinatal deaths and low birth weight (Rantakallio 1969). The original data contained 12231 (96.3%) births in the Finnish provinces of Oulu and Lapland with an expected date of birth during 1966, of which 12058 were live-born children.

Data on the cohort members were collected by self-reported questionnaires, interviews, and health examinations. Information on the socio-demographic characteristics of the mother and the family was collected at the antenatal clinic during mid-gestation (24th gestational week). The cohort members were followed-up at the ages of 6–12 months, 7–8 years, 14–16 years, and 31 years. There were also several other follow-ups for subsamples of the cohort members.

4.2 Epidemiologic health study (31-year follow-up)

The latest follow-up of all the cohort members, the 31-year follow-up, was conducted during 1997–1998 in order to study physical and mental health and their associations in the NFBC 1966 (Sorri & Järvelin 1998). Altogether 11636 cohort members were alive in the beginning of 1997 (Fig. 1). A postal questionnaire was sent to those with a known address, i.e. to 11540 (51% men) cohort members, of whom 8688 (75%) returned the postal questionnaire and did not deny the use of their data (Original study I). The questionnaire contained a comprehensive set of questions on physical and mental health, the use of health services and socio-economic status. Filling in the questionnaire was estimated to take about half an hour, on average. The subjects were asked to sign an informed consent to allow the data collected on them to be used in further research and to allow additional information on them to be collected from various registers. Altogether 93 cohort members denied the use of their data, and their data were excluded.
Live-born children with an expected date of birth during 1966, $N = 12058$, 6169 (51%) boys, 5889 (49%) girls

Died before 1997, $N = 422$ (3.5%)

Cohort members alive on 1st January 1997, $N = 11636$,
5905 (51%) men, 5731 (49%) women

Address not known, $N = 96$ (0.8%)

Phase I:

A postal questionnaire was sent in 1997, $N = 11540$,
5853 (51%) men, 5687 (49%) women

Participants, $N = 8765$ (76%)
  ◦ 4208 men (72% of men)
  ◦ 4557 women (80% of women)
Non-participants, $N = 2775$ (24%)
  ◦ 1645 men (28% of men)
  ◦ 1130 women (20% of women)

Participants, $N = 8688$ (75%)
  ◦ 4166 men (71% of men)
  ◦ 4522 women (80% of women)
Non-participants, $N = 2855$ (25%)
  ◦ 1688 men (29% of men)
  ◦ 1167 women (21% of women)

Abroad in 1982, $N = 283$ (2.5%)
Abroad in 1997, $N = 108$ (0.9%)

In further statistical analyses, subjects who lived in Finland in 1982 and 1997 and who did not refuse the use of their data were included.

Participants, $N = 8297$ (79%)
  ◦ 4003 men (75% of men)
  ◦ 4294 women (83% of women)
Non-participants, $N = 2229$ (21%)
  ◦ 1362 men (25% of men)
  ◦ 867 women (17% of women)

Refused the use of data, $N = 93$

No address, $N = 96$ (0.8%)

Abroad in 1982, $N = 424$ (3.7%)
Abroad in 1997, $N = 106$ (0.9%)

Fig. 1. Data collection of phase I (postal questionnaire) in the 31-year follow-up of the NFBC 1966 (modified from Fig. 1 in original study I).
In the second phase of the study, all 8463 (52% men) cohort members living in Northern Finland or the Helsinki area on January 1st 1997, were invited to a clinical examination in a local health center or hospital, and 6023 (71%) of them participated (Fig. 2). The procedure of the clinical examination was introduced in the invitation letter. If the subject did not attend the fixed appointment, a new appointment was offered. The examination included a blood sample draw, measurements of blood pressure, lung function and fitness, and filling in an additional self-report questionnaire. It was stated in the invitation letter that participants would be informed of the test results. The examination lasted for an average of one hour per person, and participants were not paid for participating.

Fig. 2. Data collection of phase II (clinical examination) in the 31-year follow-up of the NFBC 1966 (modified from Fig. 2 in original study I).

The third phase refers to the questionnaire handed out to participants during the clinical examination (Fig. 3). Of the 6023 subjects who participated in the clinical examination, 5133 (85%) returned the questionnaire. The questionnaire consisted of statements of opinions and experiences, including several psychological subscales. The subjects were asked to fill it in at home and return it by mail in a prepaid envelope.
4.3 Psychiatric field study

The psychiatric field study was originally based on 160 people (95 men (59%)) with a history of at least one psychotic episode by the end of 1997, according to the Finnish Hospital Discharge Register (FHDR), of whom fourteen had died by the 2001 (Fig. 4), and 187 non-psychotic control subjects (Fig. 5) (Original studies II-III).

The field study consisted of MRI scans of the brain (Tanskanen et al. 2005, Ridler et al. 2006, Tanskanen et al. 2009, Tanskanen et al. in press), mental-health-related interviews (Lauronen et al. 2005), cognitive tests (Murray et al. 2006), questionnaires, and evaluation of substance use and smoking habits. The subjects were asked to sign an informed consent at the beginning of the study. The interviews were carried out and the MRI scans were taken in Oulu University Hospital, in most cases during the course of one day. The order of the different parts of the field study was not pre-defined for subjects with a psychosis. The control subjects were scanned before the tests and interviews. Completing all parts of the field study took about four hours. If a subject was not able to finish all parts of the field study during one day, he or she was offered the possibility to continue with the tests and interviews at another time.
4.3.1 Invitation and participation of subjects with a psychosis

During the field study in 1999–2001, the most recent addresses and telephone numbers of the subjects were collected using the information service of the Population Register Centre. For subjects with a psychosis, the addresses and telephone numbers were re-examined in the cases where the information was not accurate.

The subjects were invited to participate in the field study by a letter in which the procedure of the study was introduced. A fact sheet presenting information on the MRI session was included. The subjects were informed about the progress of the MRI scanning and how to prepare for it. They were advised to notify the researchers about exclusion criteria: having a cardiac pacemaker or metal implants, having claustrophobia or being pregnant. The subjects were also told about the incentives they would receive for participating. They were promised to be given the MRI image, to have their travel expenses paid, to be given a daily allowance of 25 euros, and a snack. After the invitation letter the subjects were called by a research doctor who was prepared to answer questions concerning the study. If necessary, a maximum of three letters were sent and a maximum of three telephone calls were made in order to invite subjects with a psychosis.

Altogether 146 subjects with a history of psychosis were invited (Fig. 4). Ninety-two of them (63%) agreed to participate and provided informed consent in writing, but eventually 91 of them were eligible to participate. One patient was withdrawn because of having a metal implant in the head. Eighty-one (89%) of the participating subjects had an adequate MRI scan taken. For five subjects the scan was not taken or it failed. Five subjects have been removed from the data due to a change in diagnosis (from other psychosis to organic psychosis, non-psychotic disorder or developmental disorder) after the diagnostic interview.

4.3.2 Invitation and participation of control subjects

We intended to have two gender-matched control subjects who had not had a psychotic episode according to the FHDR for each subject with a psychosis. Non-psychotic control subjects were randomly selected from the cohort members living in the city of Oulu or in the neighbouring parishes in 1997. Due to the aims of the field study, the invitation protocol of the control subjects differed from that of the subjects with a psychosis. The addresses of the controls were searched but not re-examined. The controls received only one invitation letter. If they did not
respond to the letter or if the appointment was not suitable, another gender-matched cohort member was invited instead.

Originally 199 controls (119 men (60%)) were invited (Fig. 5). However, due to the exclusion criteria (6 subjects were pregnant, 4 had changed residence, 1 had died and 1 had Multiple Sclerosis), only 187 were eligible to participate and 104 (56%) of them consented in writing to participate in the study. Two of them did not participate in the MRI scan.

**Fig. 4. Procedure of invitation and participation of 145 members of the NFBC 1966 with a history of psychosis in the psychiatric field study conducted in 1999–2001 (modified from Fig. 1 in original study II).**
4.4 Data on medication use in the NFBC 1966

The pharmacy data on purchased prescriptions in original study IV were collected from the Finnish National Prescription Register. Drugs were identified according to the Anatomical Therapeutic Chemical (ATC) classification system (Guidelines for ATC classification 1990). In the ATC classification, drugs are classified according to the organ or system on which they act and according to their chemical, pharmacological and therapeutic properties. The data used in original study IV consisted of information on the ATC codes and the date of drug purchase, but not on the doses or amount of medication. All the members of the NFBC 1966 who had purchased any reimbursed drugs during 1997 (N = 5681; 2405 men (42%)) were selected into the study (Fig. 6), and their data on the purchase of antipsychotics (N05A), antidepressants (N06A), antiepileptics (N03A), antidiabetics (A10) and beta-blocking agents (C07) were reviewed.

Self-reported data were collected by postal questionnaire in the 31-year follow-up. The subjects were asked for the name and dosage of the drugs they used currently, and whether they were prescribed by a physician. An experienced senior psychiatrist reviewed and categorized the drugs according to the ATC classification. Those subjects who returned the questionnaire by January 31st 1998 (N = 7625; 3631 men (48%)) were selected into the study (Fig. 6). Altogether
6607 (87%) of them had reported using any of the medications listed occasionally, regularly or continuously, and 3801 (58%) had named at least one currently used drug.

**A postal questionnaire was sent in 1997, \( N = 11540 \) (5853 men (51%))**

Subjects who returned the questionnaire by 31st January 1998 and who did not refuse the use of their data, \( N = 7625 \) (66%), were selected into the study
- 3631 men (59% of men)
- 3994 women (68% of women)

Altogether 6607 subjects (2897 (44%) men) had reported using any medication, and 3801 of them (1331 (35%) men) had named the drug they used.

Altogether 326 subjects reported the use of any of the following medications:
- Antipsychotics (N05A; \( N = 65 \)), antidepressants (N06A; \( N = 87 \)), antiepileptics (N03A; \( N = 55 \)), antidiabetics (A10; \( N = 57 \)), or beta-blocking agents (C07; \( N = 95 \))

**Pharmacy data were collected from the Finnish National Prescription Register of the cohort members who had purchased any reimbursed medication during 1997, \( N = 5681 \) (2405 men (42%))**

Altogether 699 subjects had purchased any of the following medications:
- N05A (\( N = 72 \)), N06A (\( N = 117 \)), N03A (\( N = 54 \)), A10 (\( N = 54 \)), or C07 (\( N = 77 \))

---

**Fig. 6. Procedure of inclusion of the members of the NFBC 1966 into the study of self-reported medication use (modified from Fig. 1 in original study IV).**

### 4.5 Collection of data in psychiatric studies of the NFBC 1966

In the psychiatric studies of the NFBC 1966, the inclusion criteria were that the subjects lived in Finland in 1982 and 1997, and that they had not refused the use of their data. Out of 93 cohort members who had denied the use of their data, 83 lived in Finland in 1982 and 1997. Inclusion was based on the supposition that a person who does not live in Finland does not have data in the FHDR. Also, a validity study on psychiatric diagnoses of the cohort members who lived in Finland in 1982 had been conducted previously (Isohanni et al. 1997, Moilanen et al. 2003). In the psychiatric studies the target population of the postal
questionnaire was therefore 10740, of whom 5968 (56%) returned the questionnaire, and of the clinical examination 8365, of whom 5960 (71%) participated in the clinical examination and 5084 (61%) returned the additional questionnaire on psychometric assessments.

4.5.1 Data on psychiatric morbidity

The Finnish Hospital Discharge Register (FHDR) was used to extract data on the psychiatric morbidity of the subjects. The FHDR covers all psychiatric and general hospitals, in-patient wards of local health centres, and private hospitals nationwide. It consists of individually identified data on diagnoses for hospitalisation and the dates of admission and discharge. The FHDR data of the cohort members and their parents are available from 1972–2000.

All hospital case notes dealing with psychiatric disorders from January 1st 1982 to December 31st 1997 were reviewed and validated for DSM-III-R criteria in order to evaluate psychiatric morbidity arising in adult life (American Psychiatric Association 1987, Isohanni et al. 1997, Moilanen et al. 2003). Information on psychotic symptoms was collected from the hospital case notes using the Operational Criteria Checklist for Psychotic Illness (McGuffin et al. 1991). Age at the onset of illness was ascertained from the hospital case notes and defined as the year when the first psychotic symptoms were evident (Räsänen et al. 1999).

The diagnoses in the studies of this thesis were categorised according to DSM-III-R in the following way: Psychotic disorders, i.e. schizophrenia (codes 295 (except 295.40, 295.70); N = 86) and other psychoses (codes 295.40, 295.70, 296.24, 296.34, 296.44, 296.54, 296.64, 297-298; N = 51); and non-psychotic psychiatric disorders (codes 296 (except 296.24, 296.34, 296.44, 296.54, 296.64); 300-316; N = 301). Substance use disorders and mood disorders were considered separately and independently in study I as follows: substance use disorders (codes 303-305; N = 141) and mood disorders (codes 296; N = 71; including major depressive episodes (N = 53) and bipolar disorder episodes (N = 14)).

The corresponding codes according to the International Statistical Classification of Diseases and Related Health Problems 10th edition (ICD-10) are as follows: Schizophrenia (code F20), other psychoses (F22-F29, F302, F312, F315, F323, F333), non-psychotic psychiatric disorders (codes F10-F19 for substance use disorders, F30-F39 for mood disorders; and F21, F40-F69, F80-F99).
4.5.2 Other data used in this thesis

Socio-demographic information, such as educational level at the end of 1997 and marital status at the end of 1994 were collected from Statistics Finland. Information on marital status in original study IV was collected from the postal questionnaire in the 31-year follow-up study. The data on disability pensions and employment of the cohort members were collected from the Finnish Centre for Pensions (Miettunen et al. 2007).

4.5.3 Variables used in the original studies

Original study I

- **Diagnoses of psychiatric disorders.** No psychiatric disorder, any psychiatric disorder, non-psychotic disorder and psychotic disorder; specifically schizophrenia, substance use disorder, and mood disorder (see the DSM-III-R and corresponding ICD-10 codes in section 4.5.1 of this thesis)
- **Gender.** Men vs. women
- **Educational level.** Basic (9 years or less), secondary (10 to 12 years), and tertiary (over 12 years) level of education at the end of 1997 (Isohanni et al. 2001)

Original study II

- **Previous participation.** Participated vs. did not participate in the 31-year follow-up study

**Demographic factors:**

- **Gender.** Men vs. women
- **Geographic location.** Oulu and the surroundings vs. other parts of the province of Oulu, the province of Lapland, and other parts of Finland or emigrated

**Socio-economic factors:**

- **Paternal social class.** Managerial vs. manual workers and farmers in 1980 (Mäkikyrö et al. 1997)
- Smoking. Never tried or tried smoking vs. smoked occasionally or daily in 1980 (Isohanni et al. 1993)
- Marital status. Married vs. divorced or not married at the end of 1994
- Educational level. Secondary (10 to 12 years) or tertiary (over 12 years) vs. basic (9 years or less) level of education at the end of 1997 (Isohanni et al. 2001)
- Unemployment periods. Having vs. not having been unemployed during 1999–2000 (Miettunen et al. 2007)

Illness-related factors:
- Diagnosis of schizophrenia. Schizophrenia vs. non-schizophrenic psychoses by the end of 1997
- Substance use disorder. Having vs. not having substance use disorder (Räsänen et al. 1999)
- Disability pension due to a psychosis. Being vs. not being on a disability pension due to a psychosis at the end of 2000 (Miettunen et al. 2007)
- Psychotic symptoms. Presence of manic, positive, depressive and negative symptoms vs. absence of these symptoms (McGuffin et al. 1991; grouped according to Matsuura et al. (2004))
- Age at the onset of a psychosis. (Räsänen et al. 1999)
- Psychiatric hospitalisations. Cumulative days and number of treatment episodes of psychiatric hospitalisations due to a psychosis during 1972–2000
- Proportion spent in psychiatric treatment after the onset of a psychosis. Proportion calculated from the first psychotic episode until the end of 2000

Original study III

Brain volumes
- Volumes of grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). (Tanskanen et al. 2005, Ridler et al. 2006, Tanskanen et al. 2009, Tanskanen et al. in press)

Auxiliary variables:
- Disability pension due to a psychosis. Being vs. not being on a disability pension due to a psychosis at the end of 2000 (Miettunen et al. 2007)
- **Use of antipsychotic medication in 1997.** Reimbursed physician-prescribed antipsychotics purchased vs. not purchased during 1997 (Original study IV)

- **Psychiatric hospitalisations.** Cumulative days of psychiatric hospitalisations due to a psychosis during 1972–2000. Categorised into three groups (less than 75 days, 76–340 days, and over 340 days) based on the tertiles of participants

- **Age at the onset of a psychosis.** Categorised into three groups (16–19 years, 20–25 years, and 26–31 years). Used non-categorised in calculating the weights.

**Adjusting variables:**

In order to have comparable analyses with previous work (Tanskanen et al. 2009) the following adjusting variables were used:

- **Intracranial volume (ICV).** Volumes of GM, WM and CSF summed

- **Family history of psychosis.** Either of the study subject’s parents having vs. not having had a psychotic episode during 1972–2000

- **Perinatal risk.** Defined as one or more of the following events: low birth weight (< 2500 g), short gestation (< 37 weeks) or perinatal brain damage (Jones et al. 1998)

**Original study IV**

In original study IV, antipsychotics and antidepressants were selected as major psychoactive medications; and antiepileptics, antidiabetics, and beta-blocking agents as comparators for self-reporting activity.

**Medication use:**

- **Antipsychotics (ATC code: N05A), antidepressants (N06A), antiepileptics (N03A), antidiabetics (A19), and beta-blocking agents (C07).** Purchase of these medications according to the Finnish National Prescription Register, and self-reported information on the use of these medications collected by postal questionnaire in the 31-year follow-up study.

**Socio-economic data:**

- **Marital status.** Married or cohabiting vs. single, divorced or widowed in 1997
– *Educational level.* Secondary (10 to 12 years) or tertiary (over 12 years) vs. basic (9 years or less) level of education at the end of 1997 (Isohanni et al. 2001)

4.6 Statistical methods

*Original study I*

In original study I, the proportions of participants in different groups to be compared were calculated. The precision of these proportions was reported using 95% confidence intervals. Additionally, the ratio \( r_{Pr} \) between the participation rates in phases I and III was calculated:

\[
r_{Pr} = \frac{Pr(III)}{Pr(I)}
\]

where \( Pr(III) \) is the participation rate in phase III and \( Pr(I) \) in phase I. SPSS 13.0 was used to perform the statistical analyses.

*Original study II*

In original study II, for categorical variables the proportions of participants, and for continuous variables the medians, in the different groups to be compared were calculated. The differences between participants and non-participants were analysed with the chi-square test or Fisher’s exact test, and with Mann-Whitney’s test. SPSS 12.0 was used to perform the statistical analyses.

*Original study III*

Weighting was used to correct the effect of non-participation in original study III. The method takes into consideration the factors that may differ between non-participants and participants. Calculating the weights and using them in the analyses may change the estimates, leading to results of the statistical tests also changing from the results of the unweighted analyses.

The study first evaluated which variables would influence both participation and the dependent variables, i.e. brain volumes. Participation rates among subjects with schizophrenia were calculated for each auxiliary variable and
compared using Pearson’s $\chi^2$ test. The mean volumes of grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) were compared using Student’s $t$ test between the two groups of disability pension due to a psychosis and use of antipsychotic medication; and one-way analysis of variance (ANOVA) between the three groups of treatment days due to psychosis and onset age.

The weights were first calculated separately for all the auxiliary variables, and then for sets of auxiliary variables (multivariate weighting). Propensity to participate was estimated using logistic regression analysis with participation as a dependent variable and auxiliary variables as independent variables. Non-categorised onset age was used in these analyses. Weights were calculated only for subjects with schizophrenia, those for the controls being set at unity.

The volumes of GM, WM and CSF were compared between the schizophrenia and control groups using Student’s $t$ test (unadjusted analyses) and univariate ANOVA adjusted for intracranial volume (ICV), family history of psychosis and perinatal risk. All the analyses were run both unweighted and weighted, i.e. with the weighting adjustment factor $w_i$ attached to all the participating schizophrenia subjects $i$. It is the data that are weighted, so that no variable transformations are needed. In practice, e.g. in SPSS, one conducts the statistical tests as they are with the exception that the weighting procedure is activated before running the tests. SPSS 16.0 was used to perform the statistical analyses.

Original study IV

For original study IV, the prevalence of medication use based on self-reported and pharmacy data were evaluated separately. Agreement between the self-reported and pharmacy data was evaluated using Cohen’s Kappa ($\kappa$) (Cohen 1960), prevalence-adjusted and bias-adjusted Kappa (PABAK) (Byrt et al. 1993), and the proportion of positive agreement ($p_{pos}$) (Fleiss 1981). The prevalence index ($PI$) and bias index ($BI$) were also calculated to assess the distribution of medication use (prevalence) and systematic differences (bias) between the two data sets (Byrt et al. 1993).

Logistic regression was used to evaluate the association between socio-economic factors and congruence between the two data sets. Congruence was coded as: 0 = Both data sets indicated the presence of medication use, or both data sets indicated the absence of medication use; and 1 = Only one of the sources indicated the presence of medication use.
The odds ratio (OR) can be calculated using the notations of Table 2, where '+' indicates the presence of medication use, and '-' the absence of medication use in the two data sets. The letters $a-d$ denote the actual numbers of units in each cell and $f_1, f_2, g_1$, and $g_2$ the marginal totals.

The odds ratio is the cross-product ratio of the entries, i.e.

$$\hat{OR} = \frac{a \times d}{c \times b}.$$

Standard error for the natural logarithm of $\hat{OR}$ can be calculated by

$$SE(\ln \hat{OR}) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

and the 95% confidence interval for $\hat{OR}$ by

$$\exp\{\ln \hat{OR} \pm 1.96 \times SE(\ln \hat{OR})\}.$$

The effect of participation was analysed by comparing relative risks (or risk ratios; $RR$) and their 95% confidence intervals between participants and non-participants. The notations of Table 2 were used, where '+' as a column title indicates participation and '-' non-participation, and as row titles they indicate the presence or absence of medication, respectfuely.

Let $\hat{p}_1$ represent the risk estimator for the non-participation group and $\hat{p}_0$ represent the risk estimator for the participation group as follows:

$$\hat{p}_1 = \frac{a}{f_1} \quad \text{and} \quad \hat{p}_0 = \frac{c}{f_2}.$$

Relative risk can be calculated by

$$\hat{RR} = \frac{\hat{p}_1}{\hat{p}_0} = \frac{a \times f_2}{f_1 \times c}.$$

Standard error for the natural logarithm of $\hat{RR}$ can be calculated by

$$SE(\ln \hat{RR}) = \sqrt{\frac{1 - \hat{p}_1}{f_1 \times \hat{p}_1} + \frac{1 - \hat{p}_0}{f_0 \times \hat{p}_0}}$$

and the 95% confidence interval for $\hat{RR}$ by

$$\exp\{\ln \hat{RR} \pm 1.96 \times SE(\ln \hat{RR})\}.$$
WINPEPI (PEPI-for-Windows) was used to calculate the agreement statistics and the relative risks (Abramson 2004). SPSS 15.0 was used to perform other statistical analyses.

The subjects may have purchased a drug inside or outside the accepted six-month time window. The latter stands for a purchase happening either after filling in the questionnaire or more than six months prior to it. In such a case, a positive value in the self-reported data was defined as false, and negative as correct. The earlier of the dates of informed consent or a clinical examination was used as the date of filling in the questionnaire.

4.7 Ethical considerations and personal involvement

The overall study plan of the Northern Finland 1966 and 1986 Birth Cohorts was accepted by the Ethical Committee of the Northern Ostrobothnia Hospital District on February 27th 2003. The research plans were accepted by the Ethical Committee of Oulu University, Faculty of Medicine for the 31-year follow-up of the NFBC 1966 study on June 7th 1996, and for the 34-year psychiatric follow-up on March 30th 1998. Data protection was scrutinized by the Privacy Protection Agency, and according to the principles of the Ministry of Health and Social Affairs in 1994, when the permission to have information on the study sample was obtained.

Informed consent was inquired from all the participants. Subjects who denied the use of their data were excluded from the study. In this study, participants were compared with non-participants who had not signed the informed consent. According to the Personal Data Act (Personal Data Act 1999), however, personal data may be processed for purposes of scientific research on the terms presented in Chapter 2.3.2 of this thesis.

The author of this thesis has participated in the NFBC 1966 study as a researcher since the year 2003. The author has been responsible for selecting the statistical methods, participated in designing all the studies and participated in presenting the methods and results in all the original articles in this thesis in collaboration with the research group. The statistical analyses and the literature searches were done by the author. The contribution of the author in all the original articles has been central. The author has written the first and final versions of the original articles.
5 Results

5.1 Original study I

5.1.1 Effect of psychiatric morbidity on participation rates

The prevalence of any hospital-treated psychiatric disorder was 5.0% for men and 2.8% for women among all the members of the NFBC 1966 ($N = 11636$). The corresponding figures for psychotic disorders, in particular, were 1.5% and 1.2%, respectively.

Subjects with a psychiatric disorder participated less actively than those without one in all phases among both genders (Table 3), and women with a psychotic disorder participated less actively than women with a non-psychotic psychiatric disorder. However, among both men and women, the differences in participation rates between the small groups representing various types of psychiatric disorders were well within the usual error margins.

5.1.2 Effect of gender and education on participation rates

Women participated more actively than men in all phases (Table 4). Among subjects without a psychiatric disorder, a higher educational level was positively associated with participation in all phases. Men with a tertiary education participated more actively than men with a secondary education, and men with a secondary education, more actively than men with a basic education. Among subjects with a psychiatric disorder, subjects with a secondary level of education had the highest participation rates, but the differences were within the error margins. On all educational levels, however, the participation rate was lower among subjects with a psychiatric disorder.
Table 3. Proportions of the members of the NFBC 1966 participating in the 31-year follow-up study with respect to psychiatric morbidity (modified from Table I in original study I).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Psychiatric disorder</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
<th>( r ) ( \text{Pr} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N1 Pr (95% CI)</td>
<td>N2 Pr (95% CI)</td>
<td>Pr (95% CI)</td>
<td>(%)</td>
</tr>
<tr>
<td>Men</td>
<td>No psychiatric disorder</td>
<td>5085 76 (75-77)</td>
<td>4108 67 (66-68)</td>
<td>54 (52-56)</td>
<td>72.0</td>
</tr>
<tr>
<td></td>
<td>Any psychiatric disorder</td>
<td>280 58 (52-64)</td>
<td>229 46 (39-53)</td>
<td>35 (29-41)</td>
<td>60.3</td>
</tr>
<tr>
<td></td>
<td>Non-psychotic disorder</td>
<td>206 56 (49-63)</td>
<td>164 46 (38-54)</td>
<td>33 (26-40)</td>
<td>58.9</td>
</tr>
<tr>
<td></td>
<td>Psychotic disorder</td>
<td>74 62 (51-73)</td>
<td>65 48 (36-60)</td>
<td>39 (27-51)</td>
<td>62.9</td>
</tr>
<tr>
<td></td>
<td>Schizophrenia</td>
<td>52 62 (49-75)</td>
<td>44 46 (31-61)</td>
<td>39 (25-53)</td>
<td>62.9</td>
</tr>
<tr>
<td></td>
<td>Substance use disorder(^2)</td>
<td>110 52 (43-61)</td>
<td>90 40 (30-50)</td>
<td>27 (18-36)</td>
<td>51.9</td>
</tr>
<tr>
<td></td>
<td>Mood disorder(^2)</td>
<td>40 53 (37-69)</td>
<td>31 52 (34-70)</td>
<td>42 (25-59)</td>
<td>79.2</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>5365 75 (74-76)</td>
<td>4337 66 (65-67)</td>
<td>53 (51-55)</td>
<td>70.7</td>
</tr>
<tr>
<td>Women</td>
<td>No psychiatric disorder</td>
<td>5003 84 (83-85)</td>
<td>3906 77 (76-78)</td>
<td>69 (68-70)</td>
<td>82.1</td>
</tr>
<tr>
<td></td>
<td>Any psychiatric disorder</td>
<td>158 67 (60-74)</td>
<td>122 64 (55-73)</td>
<td>57 (48-66)</td>
<td>85.1</td>
</tr>
<tr>
<td></td>
<td>Non-psychotic disorder</td>
<td>95 72 (63-81)</td>
<td>67 70 (59-81)</td>
<td>63 (51-75)</td>
<td>87.5</td>
</tr>
<tr>
<td></td>
<td>Psychotic disorder</td>
<td>63 59 (47-71)</td>
<td>55 56 (43-69)</td>
<td>51 (38-64)</td>
<td>86.4</td>
</tr>
<tr>
<td></td>
<td>Schizophrenia</td>
<td>34 62 (46-78)</td>
<td>30 57 (39-75)</td>
<td>50 (32-68)</td>
<td>80.6</td>
</tr>
<tr>
<td></td>
<td>Substance use disorder(^2)</td>
<td>31 65 (48-82)</td>
<td>24 63 (44-82)</td>
<td>50 (30-70)</td>
<td>76.9</td>
</tr>
<tr>
<td></td>
<td>Mood disorder(^2)</td>
<td>31 68 (51-85)</td>
<td>22 73 (54-92)</td>
<td>68 (48-88)</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>5161 83 (82-84)</td>
<td>4028 77 (76-78)</td>
<td>69 (68-70)</td>
<td>83.1</td>
</tr>
</tbody>
</table>

Phases: I = General postal questionnaire. II = Clinical examination. III = Psychometric assessments.

N1 = Total population in Phase I. N2 = Total population in Phases II and III.

Pr (95% CI) = Proportion (95% confidence interval) of the subjects who participated in the study.

\(^1\) Calculated for comparison of proportions in phases I and III: \( r = \text{Pr}_{\text{phase II}} / \text{Pr}_{\text{phase I}} \).

\(^2\) Substance use and mood disorders are considered separately and independently.
Table 4. Proportions of the members of the NFBC 1966 participating in the 31-year follow-up study with respect to gender and education (modified from Table II in original study I).

<table>
<thead>
<tr>
<th>Psychiatric disorder</th>
<th>Gender</th>
<th>Education</th>
<th>N1</th>
<th>Phase I Pr (95% CI)</th>
<th>N2</th>
<th>Phase II Pr (95% CI)</th>
<th>N3</th>
<th>Phase III Pr (95% CI)</th>
<th>( r_{Pr} ) (^1) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Men</td>
<td>Tertiary</td>
<td>1267</td>
<td>82 (80-84)</td>
<td>937</td>
<td>77 (74-80)</td>
<td>64</td>
<td>61-67</td>
<td>78.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary</td>
<td>3092</td>
<td>76 (74-78)</td>
<td>2559</td>
<td>67 (65-69)</td>
<td>54</td>
<td>52-56</td>
<td>71.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Basic</td>
<td>718</td>
<td>64 (60-68)</td>
<td>607</td>
<td>52 (48-56)</td>
<td>38</td>
<td>34-42</td>
<td>59.4</td>
</tr>
<tr>
<td>Women</td>
<td>Tertiary</td>
<td>1411</td>
<td>88 (86-90)</td>
<td>1005</td>
<td>81 (79-83)</td>
<td>73</td>
<td>70-76</td>
<td>83.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>3118</td>
<td>85 (84-86)</td>
<td>2531</td>
<td>79 (77-81)</td>
<td>71</td>
<td>69-73</td>
<td>83.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Basic</td>
<td>471</td>
<td>65 (61-69)</td>
<td>369</td>
<td>58 (53-63)</td>
<td>50</td>
<td>45-55</td>
<td>76.9</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Men</td>
<td>Tertiary</td>
<td>15</td>
<td>53 (28-78)</td>
<td>14</td>
<td>43 (17-69)</td>
<td>29</td>
<td>5-53</td>
<td>54.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary</td>
<td>173</td>
<td>62 (55-69)</td>
<td>138</td>
<td>53 (45-61)</td>
<td>40</td>
<td>32-48</td>
<td>64.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Basic</td>
<td>92</td>
<td>50 (40-60)</td>
<td>77</td>
<td>35 (24-46)</td>
<td>26</td>
<td>16-36</td>
<td>52.0</td>
</tr>
<tr>
<td>Women</td>
<td>Tertiary</td>
<td>22</td>
<td>64 (44-84)</td>
<td>17</td>
<td>65 (42-88)</td>
<td>59</td>
<td>36-82</td>
<td>92.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>103</td>
<td>69 (60-78)</td>
<td>79</td>
<td>66 (55-77)</td>
<td>61</td>
<td>50-72</td>
<td>88.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Basic</td>
<td>33</td>
<td>61 (44-78)</td>
<td>26</td>
<td>58 (39-77)</td>
<td>46</td>
<td>27-65</td>
<td>75.4</td>
<td></td>
</tr>
</tbody>
</table>

Phases: I = Postal questionnaire. II = Clinical examination. III = Psychometric assessments.

N1 = Total population in Phase I. N2 = Total population in Phases II and III.

Pr (95% CI) = Proportion (95% confidence interval) of the subjects who participated in the study.

\(^1\) Calculated for comparison of proportions in phases I and III: \( r_{Pr} = \frac{Pr_{\text{Phase III}}}{Pr_{\text{Phase I}}} \)

5.1.3 Effect of the data collection method on participation rates

Of the cohort members eligible in phases II and III, 80% returned the postal questionnaire, 71% attended the clinical examination, and 61% returned the additional questionnaire with psychometric assessments (Figures 1–3). Among those with a psychiatric disorder, the corresponding participation rates were 64%, 52%, and 43%. Of those who participated in phase II, 85% returned the additional questionnaire on psychometric assessments.

The decline in participation rates from phase I to III was steeper among men than women, i.e. the ratio \( r_{Pr} \) was higher among women than men, irrespective of psychiatric disorder (Table 3). Among subjects without a psychiatric disorder, the ratio was higher among subjects with a tertiary or secondary level of education than among those with a basic level of education (Table 4). Among subjects with a psychiatric disorder, men with a secondary level of education had a higher ratio than men with a tertiary or basic level of education, whereas among women the ratio was positively associated with the level of education.
5.2 Original study II

5.2.1 Non-participation among subjects with a psychosis

A total of 91 (63%) out of 145 subjects with a psychosis attended the psychiatric field study (Fig. 4). Of the 54 non-participating subjects, three (6%) were not contacted, 30 (56%) refused to participate, three (6%) did not arrive to the appointment and 18 (32%) did not respond to the invitation letters or the calls made by telephone.

Participation in the 31-year follow-up study was highly associated with participation in this study (Table 5). Those who lived nearby and within easy transportation to the research centre participated more commonly, although the differences were not significant. Socio-economic factors in youth and adulthood did not differ between participants and non-participants, except for participants more often being married.

Subjects with schizophrenia participated less actively than those with other psychoses (Table 6). Those on a disability pension due to a psychosis and with more positive symptoms were more likely to be non-participants. Non-participants had had more psychiatric hospitalisations than participants (Median 6 vs. 4 times, \( P = 0.046 \); and 248 vs. 113 days, \( P = 0.076 \)). Age at onset of illness did not differ between non-participants and participants (23 vs. 24 years, \( P = 0.354 \)).

When ten subjects without an adequate MRI scan were excluded and all the analyses were repeated, the results were essentially the same (detailed data not shown), but diagnosis (schizophrenia vs. other psychoses) did not reach statistical significance (\( P = 0.388 \)). Also, those who refused and who were not contacted did not differ except that the non-contacted were more often men.
Table 5. Demographic and socioeconomic factors with respect to participation in the psychiatric field study in 1999–2001 (modified from Table 1 in original study II).

<table>
<thead>
<tr>
<th></th>
<th>Participants (N = 91)</th>
<th>Non-participants (N = 54)</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participation in the 31-year follow-up study</td>
<td></td>
<td></td>
<td>0.001&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Yes</td>
<td>87 64 73.6</td>
<td>23 26.4</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>58 27 46.6</td>
<td>31 53.4</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td>1.000&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Male</td>
<td>83 52 62.7</td>
<td>31 37.3</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>62 39 62.9</td>
<td>23 37.1</td>
<td></td>
</tr>
<tr>
<td>Living area</td>
<td></td>
<td></td>
<td>0.202&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Oulu and the surroundings</td>
<td>27 20 74.1</td>
<td>7 25.9</td>
<td></td>
</tr>
<tr>
<td>Rest of the province of Oulu</td>
<td>56 38 67.9</td>
<td>18 32.1</td>
<td></td>
</tr>
<tr>
<td>The province of Lapland</td>
<td>27 15 55.6</td>
<td>12 44.4</td>
<td></td>
</tr>
<tr>
<td>Rest of Finland or emigrated</td>
<td>35 18 51.4</td>
<td>17 48.6</td>
<td></td>
</tr>
<tr>
<td>At or by the age of 14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paternal social class</td>
<td></td>
<td></td>
<td>0.324&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Managerials</td>
<td>41 24 58.5</td>
<td>17 41.5</td>
<td></td>
</tr>
<tr>
<td>Manual workers</td>
<td>89 55 61.8</td>
<td>34 38.2</td>
<td></td>
</tr>
<tr>
<td>Farmers</td>
<td>15 12 80.0</td>
<td>3 20.0</td>
<td></td>
</tr>
<tr>
<td>Family type</td>
<td></td>
<td></td>
<td>0.692&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Two-parent family</td>
<td>109 67 61.5</td>
<td>42 38.5</td>
<td></td>
</tr>
<tr>
<td>Single-parent family</td>
<td>36 24 66.7</td>
<td>12 33.3</td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td>0.480&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Has never or has tried smoking</td>
<td>95 58 61.1</td>
<td>37 38.9</td>
<td></td>
</tr>
<tr>
<td>Smokes occasionally or daily</td>
<td>24 17 70.8</td>
<td>7 29.2</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>26 16 61.5</td>
<td>10 38.5</td>
<td></td>
</tr>
<tr>
<td>In adulthood</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital status in the end of 1994</td>
<td></td>
<td></td>
<td>0.010&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Married</td>
<td>21 19 90.5</td>
<td>2 9.5</td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td>5 4 80.0</td>
<td>1 20.0</td>
<td></td>
</tr>
<tr>
<td>Not married</td>
<td>119 68 57.1</td>
<td>51 42.9</td>
<td></td>
</tr>
<tr>
<td>Educational level in the end of 1997</td>
<td></td>
<td></td>
<td>0.538&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Basic</td>
<td>29 16 55.2</td>
<td>13 44.8</td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>107 70 65.4</td>
<td>37 34.6</td>
<td></td>
</tr>
<tr>
<td>Tertiary</td>
<td>9 5 55.6</td>
<td>4 44.4</td>
<td></td>
</tr>
<tr>
<td>Unemployment periods since 1998</td>
<td></td>
<td></td>
<td>0.845&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>No</td>
<td>108 67 62.0</td>
<td>41 38.0</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>37 24 64.9</td>
<td>13 35.1</td>
<td></td>
</tr>
</tbody>
</table>

Significance from<sup>1</sup> Fisher’s exact two-sided test, and<sup>2</sup> Pearson’s two-sided $\chi^2$ test.
Table 6. Illness-related factors with respect to participation in the psychiatric field study in 1999–2001 (modified from Table 2 in original study II).

<table>
<thead>
<tr>
<th></th>
<th>Participants (N = 91)</th>
<th>Non-participants (N = 54)</th>
<th>Sig.¹</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>N %</td>
<td>N %</td>
</tr>
<tr>
<td><strong>Diagnosis by the end of 1997</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSM-III-R schizophrenia</td>
<td>91</td>
<td>51 56.0</td>
<td>40 44.0</td>
</tr>
<tr>
<td>Other psychosis</td>
<td>54</td>
<td>40 74.1</td>
<td>14 25.9</td>
</tr>
<tr>
<td><strong>Substance abuse</strong></td>
<td></td>
<td></td>
<td>⁰⁴⁶³</td>
</tr>
<tr>
<td>No</td>
<td>125</td>
<td>80 64.0</td>
<td>45 36.0</td>
</tr>
<tr>
<td>Yes</td>
<td>20</td>
<td>11 55.0</td>
<td>9 45.0</td>
</tr>
<tr>
<td><strong>On disability pension due to psychosis</strong></td>
<td></td>
<td></td>
<td>⁰⁰¹⁶</td>
</tr>
<tr>
<td>No</td>
<td>68</td>
<td>50 73.5</td>
<td>18 26.5</td>
</tr>
<tr>
<td>Yes</td>
<td>77</td>
<td>41 58.4</td>
<td>36 46.6</td>
</tr>
<tr>
<td><strong>Psychotic symptoms:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manic symptoms²</td>
<td></td>
<td></td>
<td>⁰²⁸⁰</td>
</tr>
<tr>
<td>Not present</td>
<td>88</td>
<td>50 79.5</td>
<td>38 20.5</td>
</tr>
<tr>
<td>Present</td>
<td>30</td>
<td>21 70.0</td>
<td>9 30.0</td>
</tr>
<tr>
<td>Positive symptoms²</td>
<td></td>
<td></td>
<td>⁰⁰³³</td>
</tr>
<tr>
<td>Not present</td>
<td>35</td>
<td>28 80.0</td>
<td>7 20.0</td>
</tr>
<tr>
<td>Present</td>
<td>46</td>
<td>26 56.5</td>
<td>20 43.5</td>
</tr>
<tr>
<td>Depressive symptoms²</td>
<td></td>
<td></td>
<td>⁰³⁸⁵</td>
</tr>
<tr>
<td>Not present</td>
<td>51</td>
<td>27 52.9</td>
<td>24 47.1</td>
</tr>
<tr>
<td>Present</td>
<td>35</td>
<td>22 62.9</td>
<td>13 37.1</td>
</tr>
<tr>
<td>Negative symptoms²</td>
<td></td>
<td></td>
<td>⁰³⁰⁹</td>
</tr>
<tr>
<td>Not present</td>
<td>66</td>
<td>43 65.2</td>
<td>23 34.8</td>
</tr>
<tr>
<td>Present</td>
<td>41</td>
<td>22 53.7</td>
<td>19 46.3</td>
</tr>
</tbody>
</table>

¹ Significance from Fisher’s exact two-sided test. ² Matsuura et al. (2004).

5.2.2 Non-participation among control subjects

A total of 104 (56%) control subjects participated in the psychiatric field study (Fig. 5). Of the non-participating controls, 32 (40%) refused to participate, 15 (18%) could or did not arrive to the appointment, 16 (19%) did not respond to the invitation letter and 20 (23%) did not participate for some unspecified reason.

We analysed the same demographic and socioeconomic factors for the controls as we did for subjects with a psychosis (detailed data not shown). Due to
the selection criteria, all the controls were from the city of Oulu or nearby. Participants differed from non-participants only in participating more actively also in the 31-year follow-up study (89% vs. 71%; \( P = 0.005 \)).

5.3 Original study III

Altogether 101 out of 145 subjects with a psychosis had schizophrenia, and 61 (60%) of them participated in the psychiatric field study (Fig. 7). Out of 101 subjects with schizophrenia and 187 non-psychotic control subjects, 54 (54%) and 100 (54%), respectively, had an adequate MRI scan of GM, WM and CSF taken.

![Diagram](image_url)

**Fig. 7. Procedure of invitation and participation of 101 subjects with schizophrenia and 187 non-psychotic control subjects in the psychiatric field study including an MRI scan of the brain (modified from Fig. 1 in original study III).**
5.3.1 Non-participation among subjects with schizophrenia

Subjects who were receiving a disability pension due to a psychosis participated less actively than those who were not (Table 7). There were no statistically significant differences in participation related to the other auxiliary variables but participants tended to have been prescribed antipsychotic medication more often, to have had less treatment days due to a psychosis, and to have had a later onset of their illness.

Table 7. Participation with respect to auxiliary variables among subjects with schizophrenia in the psychiatric field study in 1999–2001 (Table 1 in original study III).

<table>
<thead>
<tr>
<th>Auxiliary variables</th>
<th>Non-participants (N = 47)</th>
<th>Participants (N = 54)</th>
<th>P²</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>At disability pension¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>36</td>
<td>10 27.8</td>
<td>26</td>
<td>72.2</td>
</tr>
<tr>
<td>Yes</td>
<td>65</td>
<td>37 56.9</td>
<td>28</td>
<td>43.1</td>
</tr>
<tr>
<td>Use of antipsychotics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>66</td>
<td>35 53.0</td>
<td>31</td>
<td>47.0</td>
</tr>
<tr>
<td>Yes</td>
<td>35</td>
<td>12 34.3</td>
<td>23</td>
<td>65.7</td>
</tr>
<tr>
<td>Treatment days²</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 75</td>
<td>28</td>
<td>10 35.7</td>
<td>18</td>
<td>64.3</td>
</tr>
<tr>
<td>76 to 340</td>
<td>32</td>
<td>14 43.8</td>
<td>18</td>
<td>56.2</td>
</tr>
<tr>
<td>Over 340</td>
<td>41</td>
<td>23 56.1</td>
<td>18</td>
<td>43.9</td>
</tr>
<tr>
<td>Onset age⁴ (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 to 19</td>
<td>29</td>
<td>15 51.7</td>
<td>14</td>
<td>48.3</td>
</tr>
<tr>
<td>20 to 25</td>
<td>43</td>
<td>21 48.8</td>
<td>22</td>
<td>51.2</td>
</tr>
<tr>
<td>26 to 31</td>
<td>29</td>
<td>11 37.9</td>
<td>18</td>
<td>62.1</td>
</tr>
</tbody>
</table>

¹ Those with a succeeded MRI scan. ² Due to psychosis. ³ Significance from the χ² test.
⁴ Onset age was used as continuous in later analyses.

5.3.2 Brain volume estimates in subjects with schizophrenia

There were no significant differences in the volumes of GM and WM related to any of the auxiliary variables (Table 8), although those who were on a disability pension due to a psychosis had somewhat more CSF than those who were not.

The mean volumes of GM, WM and CSF in the control group were 697 ml, 600 ml and 184 ml; respectively (Table 9), and the corresponding unweighted
mean volumes among the subjects with schizophrenia were 682 ml ($P = 0.136$), 584 ml ($P = 0.146$) and 196 ml ($P = 0.066$).

| Table 8. Mean volumes (ml) of GM, WM and CSF with respect to auxiliary variables among subjects with schizophrenia in the psychiatric field study in 1999–2001 (Table 2 in original study III). |
|---|---|---|---|---|---|---|
| Auxilliary variables | Total | Mean (SD) | $P^2$ | Mean (SD) | $P^2$ | Mean (SD) | $P^2$ |
| **At disability pension**$^1$ | | | | | | | |
| No | 26 | 685.5 (50.1) | 584.2 (58.4) | 183.0 (35.3) | 0.712 | 0.954 | 0.024 |
| Yes | 28 | 679.6 (64.8) | 583.1 (75.9) | 208.1 (43.2) | 0.583 | 0.925 | 0.558 |
| **Use of antipsychotics** | | | | | | | |
| No | 31 | 678.7 (61.9) | 582.8 (61.6) | 198.9 (37.9) | 0.536 | 0.776 | 0.413 |
| Yes | 23 | 687.5 (52.5) | 584.6 (75.9) | 192.2 (45.9) |
| **Treatment days**$^2$ | | | | | | | |
| Less than 75 days | 18 | 686.6 (55.4) | 586.6 (65.8) | 185.9 (27.0) | 0.583 | 0.925 | 0.558 |
| 76 to 340 days | 18 | 690.6 (63.0) | 589.8 (81.6) | 198.3 (49.7) |
| Over 340 days | 18 | 670.1 (55.8) | 574.4 (55.1) | 203.9 (43.9) |
| **Onset age**$^3$ | | | | | | | |
| 16 to 19 years | 14 | 676.9 (77.8) | 589.7 (81.5) | 204.7 (47.3) | 0.750 | 0.727 | 0.310 |
| 20 to 25 years | 22 | 679.0 (48.9) | 588.3 (55.7) | 200.3 (43.8) |
| 26 to 31 years | 18 | 690.9 (51.8) | 573.1 (71.2) | 184.1 (31.2) |

GM = grey matter, WM = white matter, CSF = cerebrospinal fluid.

$^1$ Due to psychosis. $^2$ Significance from the Student's t test or one-way ANOVA.

5.3.3 Comparing weighted and unweighted analyses of brain volumes between the schizophrenia and control groups

Weighting the analyses of GM, WM, and CSF volumes with individual auxiliary variables slightly reduced the estimated mean volumes of GM, but did not affect those of WM, in subjects with schizophrenia (Table 9). Weighting by a disability pension and having had more treatment days due to psychosis increased the estimated mean volume of CSF and led to a statistically significant difference between the schizophrenia and control groups. The most effective combination in multivariate weighting was a disability pension and antipsychotic medication, leading to the same estimated mean volumes of GM, WM and CSF as did weighting by all the auxiliary variables together.
Table 9. Mean volumes (ml) of GM, WM and CSF among control subjects and the subjects with schizophrenia, unweighted and weighted by the auxiliary variables in a psychiatric field study in 1999–2001 (Table 3 in the original study III).

<table>
<thead>
<tr>
<th></th>
<th>GM</th>
<th>WM</th>
<th>CSF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD) P</td>
<td>Mean (SD) P</td>
<td>Mean (SD) P</td>
</tr>
<tr>
<td>Controls (n = 100)</td>
<td>696.9 (57.3)</td>
<td>599.9 (65.6)</td>
<td>184.3 (35.2)</td>
</tr>
<tr>
<td>Unweighted SZ (n = 54)</td>
<td>682.4 (57.7)</td>
<td>536.3 (67.4)</td>
<td>196.0 (41.2)</td>
</tr>
<tr>
<td>SZ weighted by</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disability pension(^1)</td>
<td>681.7 (59.5)</td>
<td>538.5 (69.5)</td>
<td>199.2 (42.0)</td>
</tr>
<tr>
<td>Use on antipsychotics</td>
<td>681.7 (58.4)</td>
<td>538.5 (66.3)</td>
<td>196.6 (40.6)</td>
</tr>
<tr>
<td>Treatment days(^1)</td>
<td>681.3 (57.7)</td>
<td>528.7 (66.6)</td>
<td>197.2 (42.0)</td>
</tr>
<tr>
<td>Onset age</td>
<td>682.1 (58.8)</td>
<td>584.1 (67.8)</td>
<td>196.7 (41.5)</td>
</tr>
<tr>
<td>Multivariate weighting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disability pension(^1) with</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of antipsychotics</td>
<td>679.3 (59.9)</td>
<td>581.2 (67.8)</td>
<td>199.7 (41.6)</td>
</tr>
<tr>
<td>Treatment days(^1)</td>
<td>681.3 (59.3)</td>
<td>538.1 (68.8)</td>
<td>199.2 (41.9)</td>
</tr>
<tr>
<td>Onset age</td>
<td>681.6 (59.2)</td>
<td>583.1 (69.4)</td>
<td>199.1 (41.9)</td>
</tr>
<tr>
<td>All auxiliary variables</td>
<td>679.3 (59.8)</td>
<td>581.1 (67.8)</td>
<td>199.7 (41.7)</td>
</tr>
</tbody>
</table>

GM = grey matter, WM = white matter, CSF = cerebrospinal fluid. SZ = schizophrenia.

\(^1\) Due to psychosis. \(^2\) Significance from the Student’s t test.

Control subjects had more GM and WM and less CSF than subjects with schizophrenia in the unweighted ANOVA when adjusted for ICV, family history of psychosis and perinatal risk (Table 10). In all the weighted analyses the estimated mean volume of GM was statistically significantly larger in the control group than in the schizophrenia group. Being on a disability pension and having more treatment days due to a psychosis had the most effect in the weighted analyses, leading to statistically significant differences in the estimated mean volumes of WM and increasing the statistical significance of the difference in the estimated mean volume of CSF between the schizophrenia and control groups.
Table 10. Adjusted effect of schizophrenia relative to the control subjects. Analysis of variance with volumes of GM, WM and CSF as individual dependent variables and diagnosis of schizophrenia as an independent variable; adjusted for intracranial volume, family history of psychosis and perinatal risk; unweighted and weighted by the auxiliary variables in the psychiatric field study in 1999–2001 (Modified from Table 4 in original study IV).

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>GM</th>
<th>WM</th>
<th>CSF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est. mean</td>
<td>Est. mean</td>
<td>Est. mean</td>
</tr>
<tr>
<td></td>
<td>SZ</td>
<td>C</td>
<td>P^4</td>
</tr>
<tr>
<td>Unweighted ANOVA</td>
<td>688.9</td>
<td>696.4</td>
<td>0.054</td>
</tr>
<tr>
<td>Schizophrenia weighted by</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disability pension^2</td>
<td>687.8</td>
<td>697.0</td>
<td>0.019</td>
</tr>
<tr>
<td>Use of antipsychotics</td>
<td>688.6</td>
<td>696.6</td>
<td>0.040</td>
</tr>
<tr>
<td>Treatment days^2</td>
<td>688.1</td>
<td>696.5</td>
<td>0.031</td>
</tr>
<tr>
<td>Onset age</td>
<td>688.6</td>
<td>696.8</td>
<td>0.037</td>
</tr>
<tr>
<td>All auxiliary variables</td>
<td>686.7</td>
<td>696.7</td>
<td>0.012</td>
</tr>
</tbody>
</table>

GM = grey matter, WM = white matter, CSF = cerebrospinal fluid.
^1 SZ = schizophrenia group, C = non-psychotic control group. ^2 Due to psychosis.
^3 Estimated means of GM, WM and CSF volumes. ^4 Significance from the analysis of variance.

5.4 Original study IV

5.4.1 Prevalence estimates of medication use according to pharmacy and self-reported data

According to the pharmacy data, 699 (6.0%) out of the 11636 cohort members had purchased at least one of the medications selected to this study during 1997, as had 326 (4.3%) out of the 7625 subjects who returned the questionnaire by January 31st 1998. The corresponding figures for the psychoactive medications (N05A or N06A) were 406 (3.5%) and 140 (1.8%).

Out of 7625 subjects in our sample (Fig. 6), 72 had purchased antipsychotics, 117 antidepressants, 54 antiepileptics, 54 antidiabetics, and 77 beta-blockers (Table 11). The corresponding figures were 65, 87, 55, 57, and 95 in the self-reported data. The proportion of subjects who had purchased medications within all the cohort members was higher than within our sample.
Table 11. Self-reported and pharmacy data on medication use in 1997 within the sample of 7625 cohort members, and pharmacy data on all 11636 cohort members, in the NFBC 1966 (Table 1 in original study IV).

<table>
<thead>
<tr>
<th>Medication</th>
<th>ATC-code</th>
<th>Responders¹</th>
<th>All cohort members²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Self-reported data</td>
<td>Pharmacy data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Antipsychotics</td>
<td>N05A</td>
<td>65 (0.9)</td>
<td>72 (0.9)</td>
</tr>
<tr>
<td>Antidepressants</td>
<td>N06A</td>
<td>87 (1.3)</td>
<td>117 (1.1)</td>
</tr>
<tr>
<td>Antiepileptics</td>
<td>N03A</td>
<td>55 (0.7)</td>
<td>54 (0.7)</td>
</tr>
<tr>
<td>Antidiabetics</td>
<td>A10</td>
<td>57 (0.7)</td>
<td>54 (0.7)</td>
</tr>
<tr>
<td>Beta-blocking agents</td>
<td>C07</td>
<td>95 (1.2)</td>
<td>77 (1.0)</td>
</tr>
</tbody>
</table>

ATC = Anatomical Therapeutic Chemical classification system. n (%) = Number (proportion).
¹ Subjects who returned the postal questionnaire by January 31st 1997 (N = 7625).
² All the cohort members alive in the beginning of January 1997 (N=11636).

5.4.2 Precision of self-reported medication use

The precision of self-reported data compared with pharmacy data was highest for antidiabetics and lowest for beta-blocking agents (Table 12). Precision was substantial for antipsychotics (κ = 0.77) and antidepressants (κ = 0.68), and almost perfect for antiepileptics (κ = 0.84).

The prevalence index was generally high for all medications (|PI| > 0.97), indicating that the two data sets had a substantial proportion of subjects without medication (Table 12). The bias index, on the other hand, was low (BI ≤ 0.004), indicating equal proportions of medication use in both data sets. Owing to the effects of prevalence and bias, the PABAK values were very high (PABAK ≥ 0.98).

Table 12. Precision of the self-reported and pharmacy data on medication use in 1997 in the NFBC 1966 (Table 2 in original study IV).

<table>
<thead>
<tr>
<th>Medication</th>
<th>Total</th>
<th>Ppos</th>
<th>κ (95% CI)</th>
<th></th>
<th>PI</th>
<th></th>
<th>BI</th>
<th>PABAK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antipsychotics</td>
<td>84</td>
<td>0.774</td>
<td>0.77 (0.69-0.85)</td>
<td>0.982</td>
<td>0.001</td>
<td>0.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antidepressants</td>
<td>134</td>
<td>0.686</td>
<td>0.68 (0.61-0.76)</td>
<td>0.973</td>
<td>0.004</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antiepileptics</td>
<td>63</td>
<td>0.844</td>
<td>0.84 (0.77-0.92)</td>
<td>0.986</td>
<td>0</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antidiabetics</td>
<td>60</td>
<td>0.919</td>
<td>0.92 (0.87-0.97)</td>
<td>0.985</td>
<td>0</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-blocking agents</td>
<td>124</td>
<td>0.558</td>
<td>0.55 (0.46-0.64)</td>
<td>0.978</td>
<td>0.002</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total refers to the subjects who reported using the particular drug or had purchased it.
Ppos = Proportion of positive agreement, κ (95% CI) = Cohen’s Kappa (95% confidence interval),
|PI| = prevalence index, BI = Bias index, PABAK = Prevalence-adjusted and bias-adjusted kappa.
5.4.3 Effect of socio-economic factors on the precision of self-reported medication use

Gender did not affect the precision of self-reported psychoactive medication use (Table 13). Men reported the use of beta-blocking agents considerably better than women. Congruence between the two data sets was especially poor among women; 27% of women underreported and 46% over-reported the use of beta-blocking agents. Both men and women reported the use of antidiabetics accurately.

Subjects with a secondary or tertiary education reported the use of antidepressants and antiepileptics more precisely than those with a basic education, and married subjects reported the use of antipsychotics and antidepressants more precisely than those who were not married. The level of education or marital status did not affect reporting of antidiabetics or beta-blocking agents.

Table 13. Effect of socio-demographic factors on discordance between the self-reported and pharmacy data on medication use in 1997 in the NFBC 1966.

<table>
<thead>
<tr>
<th></th>
<th>Gender¹</th>
<th>Education²</th>
<th>Marital status³</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>N</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>Antipsychotics</td>
<td>84</td>
<td>43 1.07 (0.52-2.21)</td>
<td>20 1.99 (0.81-4.93)</td>
</tr>
<tr>
<td>Antidepressants</td>
<td>134</td>
<td>61 0.92 (0.56-1.52)</td>
<td>24 1.95 (1.03-3.70)</td>
</tr>
<tr>
<td>Antiepileptics</td>
<td>63</td>
<td>32 1.62 (0.48-5.49)</td>
<td>19 7.67 (2.43-24.2)</td>
</tr>
<tr>
<td>Antidiabetics</td>
<td>60</td>
<td>31 0.98 (0.24-3.98)</td>
<td>5 2.91 (0.58-14.7)</td>
</tr>
<tr>
<td>Beta-blocking agents</td>
<td>124</td>
<td>54 0.53 (0.32-0.86)</td>
<td>19 0.86 (0.37-1.99)</td>
</tr>
</tbody>
</table>

¹Men vs. women. ²Basic vs. higher education. ³Not married (single, divorced or widowed) vs. married.

OR (95% CI) = Odds ratio (95% confidence interval).

5.4.4 Effect of non-participation on self-reported medication use

Non-participants, i.e. those who did not return the questionnaire by the specified time, had purchased antipsychotics and antidepressants (RR = 1.9 and RR = 1.4, respectively) more commonly than participants (Table 14). Participation did not differ with respect to other medications.
Table 14. Medication use in the NFBC 1966 with respect to participation status, i.e. subjects who returned the questionnaire by the specified time vs. subjects who did not.

<table>
<thead>
<tr>
<th>Medication</th>
<th>Participants N (%)</th>
<th>Non-participants N (%)</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antipsychotics</td>
<td>85 (1.1%)</td>
<td>85 (2.1%)</td>
<td>1.90 (1.41, 2.56)</td>
</tr>
<tr>
<td>Antidepressants</td>
<td>170 (2.2%)</td>
<td>127 (3.2%)</td>
<td>1.42 (1.13, 1.78)</td>
</tr>
<tr>
<td>Antiepileptics</td>
<td>57 (0.7%)</td>
<td>38 (0.9%)</td>
<td>1.27 (0.84, 1.91)</td>
</tr>
<tr>
<td>Antidiabetics</td>
<td>58 (0.8%)</td>
<td>36 (0.9%)</td>
<td>1.18 (0.78, 1.79)</td>
</tr>
<tr>
<td>Beta blocking agents</td>
<td>111 (1.5%)</td>
<td>63 (1.6%)</td>
<td>1.08 (0.79, 1.47)</td>
</tr>
<tr>
<td>Total</td>
<td>7625</td>
<td>4011</td>
<td></td>
</tr>
</tbody>
</table>

N (%) = Number (proportion). RR (95% CI) = Relative risk (95% confidence interval).
The pharmacy data has not been corrected by the dates of purchase and self-report in this analysis.
6 Discussion

6.1 Main findings

In original study I, in an epidemiologic study of a large population-based birth cohort, subjects with a psychiatric disorder participated less actively than those without among both genders. However, the type of disorder did not seem to affect participation. The study consisted of a self-reported postal questionnaire (phase I), a health examination (phase II) and additional psychometric assessments (phase III). Participation rates decreased from phase I to III more steeply among men than women, irrespective of the psychiatric disorder.

In original study II, in a psychiatric field study of subjects with a psychosis, non-participants were more often patients with schizophrenia than with other psychoses. The psychiatric symptoms of non-participants were more severe and they needed more hospital care than participants. Pre-morbid social or familial factors (except being married) did not predict participation.

In original study III, it was shown that inverse probability weighting (IPW) can be used to alter brain volume estimates affected by non-participation in order to better reflect true differences in the target population. Use of IPW in the schizophrenia group led to increased estimates of cerebrospinal fluid (CSF) volume in unadjusted analyses and smaller estimates of grey matter (GM) volume in adjusted analyses. Weighting by being on a disability pension and having a higher number of treatment days due to a psychosis led to smaller estimates of white matter (WM) volume in the adjusted analyses and also increased the statistical significance of the differences in CSF between the schizophrenia and control groups.

In original study IV, agreement between self-reported and pharmacy data on psychoactive medication use was substantial. Agreement was, however, generally highest in the use of antidiabetics and lowest in the use of beta-blocking agents. Gender did not affect the precision of psychoactive medication reports, and the effect of education and marital status varied within different medications.
6.2 Discussion on the results

6.2.1 Original study I

Non-participation in the 31-year follow-up study

Field studies are often used to estimate the prevalence and risk factors of psychiatric disorders. According to Drivsholm et al. (2006), subjects who participate in health studies have better health profiles than those who do not. In the population-based study by Drivsholm et al. (2006), non-participants were shown to have had episodes of psychiatric hospitalisations more often than participants. Some previous studies of psychiatric patients have suggested that more psychiatric problems occur among non-participants than among participants (Fischer et al. 2001, Vanable et al. 2002). Also Koskinen et al. (2008) stated in the methodology report of the Health 2000 study that participation in health surveys is significantly lower among people with severe mental health problems. This would lead to biased prevalence estimates of psychiatric disorders, these being lower than true prevalence by an unknown magnitude. Higher psychiatric morbidity among non-participants was also observed in the large, population-based birth cohort in the epidemiological 31-year follow-up of the NFBC 1966.

Participation may be affected by practical issues, such as work situation and residential area, as well as several psychological aspects like personality, attitude towards surveys, or passivity in general. The specific effects of all of these on participation are largely unknown. Passivity, as well as cognitive deficits, for example, may sometimes be symptoms in specific psychiatric disorders (Murray et al. 2006), although these features are not uncommon in healthy persons, either. Overlapping problems may affect participation with respect to psychiatric disorders, i.e. the issues that affect participation in general, e.g. low education, low socio-economic situation, stigma, are accentuated in the participation of subjects with psychiatric disorders. Also a number of psychiatric disorders have been associated with both refusal and loss of contact (Eaton et al. 1997).

Effect of the type of psychiatric disorder on non-participation

Insufficient evidence was found for the participation rate being dependent on the type of psychiatric disorder. This may well be due to the small size of the specific
diagnostic groups. Schizophrenia is considered to be the most severe psychiatric illness, and in some previous studies non-participation has been shown to accumulate into the group of schizophrenia patients (Bijl et al. 1998, Susser et al. 2000, Lundberg et al. 2005). However, in original study I the participation rate was at the same level for schizophrenia as for less severe disorders. There were some minor differences between different psychiatric disorders: subjects with mood disorder were slightly more active in participating, and especially men with substance use disorder were slightly more passive.

Effect of gender and education on non-participation

Women participated more actively than men and highly educated subjects more than those with a lower educational level. These observations have been made in other studies, too (Eerola et al. 2005, Lundberg et al. 2005, Carlsson et al. 2006, Koskinen et al. 2008). Among subjects with a psychiatric disorder, however, those who had a tertiary education participated less commonly than those who had a secondary education among both genders. It may be that contracting a psychiatric illness causes more dramatic consequences for highly educated subjects than for those with less education. In the NFBC 1966, good school performance was related to an increased risk for suicide among psychotic persons (Alaräisänen et al. 2006).

Subjects with lower education may have difficulties in filling in the self-reported questionnaires. Consideration should be given to approaching subjects with such difficulties with more easily understandable and less detailed questionnaires. They may also have lower health literacy, i.e. less understanding of the value of medical science.

Effect of the method of data collection on non-participation

It is presumably easier to participate in a study with data collected by postal questionnaire than in a study involving a clinical examination, purely due to practical reasons. The response rates in clinical examinations were lower than those in postal questionnaire surveys in the population-based FINRISK study on cardiovascular diseases (Peltonen et al. 2008). Also in original study I, the response rates in the clinical examination were lower than those of the postal questionnaire. Response rates decreased even further from phase I to III in all subgroups of psychiatric disorders, and more among men than women.
The 31-year follow-up study included rather personal items about the subjects’ mental health. It is quite conceivable that this might have affected the response rate. However, Peters et al. (1998) did not find that sensitive questions would influence the response rate. They emphasised that important, albeit potentially intrusive, questions should not be withheld for fear of reducing participation. According to a recent review, the response rates in epidemiologic cohort studies have declined in recent years (Morton et al. 2006), which was also seen in the FINRISK surveys conducted during 1972–1992 (Harald et al. 2007). The response rates in original study I were surprisingly good, especially among subjects with psychiatric disorders, compared with many health studies elsewhere.

The 31-year follow-up of the NFBC 1966 was an epidemiologic health study which aimed to study all the cohort members. Altogether 76% of the eligible cohort members returned the postal questionnaire, and 71% of the eligible cohort members attended the clinical examination. The participation rates were lower among subjects with psychiatric disorders. In the Health 2000 Survey altogether 84% returned the postal questionnaire and 80% attended the health examination. In these large population-based studies it may be difficult to make recruitment much more effective due to practical reasons, e.g. economic or human resources. Recruitment of special groups, e.g. subjects with psychiatric disorders, could have been emphasized in the 31-year follow-up. However, the focus of the study was not specifically on psychiatric disorders. As a consequence of original study I, recruitment of subjects who are least likely to participate could be strengthened so that the sample would be representative of all the cohort members. Subjects with e.g. low education or a psychiatric disorder should be contacted more often. Those who do not reply after several reminders should be offered an abridged questionnaire or an interview at home in order to collect the most important information.

6.2.2 Original study II

In original study II, the participation rates of patients with schizophrenia were lower than of those with other psychoses. The MRI scan of the brain was a salient part of the field study. Some of the patients were willing to take part in other parts of the field study but not the MRI scan. Some consented to attend, but did not get an adequate MRI scan after all.

The study required each participant to visit Oulu University Hospital for 4–5 hours, which may have been stressful for some of the subjects. Many psychotic
subjects may have difficulties in coping with stress and doing activities demanding complex cognitive skills. Taken together, the interviews, questionnaires, psychological and cognitive tests, as well as the MRI of the brain, required a lot from the participants, and a person could be expected to be reasonably well to be able to cope with it. In order to increase participation and compensate for participation, the travel expenses of the subjects were paid and they received a daily allowance and a snack on the examination day.

Socio-demographic factors associated with participation

In the Health 2000 Survey, the participation rate was highest among people living in the catchment area of Oulu University Hospital, and lowest in the Helsinki and Turku areas (Koskinen et al. 2008). In original study II, those who lived near the research centre, i.e. those living in the city of Oulu, had the highest participation rate. It takes a lot of time and effort to travel from other parts of Finland, especially from distant areas of Northern Finland or Eastern Finland, to the research centre. Most likely it is particularly difficult for subjects with a psychiatric illness.

Being married or cohabiting (e.g. Lundberg et al. 2005), having a high level of education (e.g. Eerola et al. 2005, Lundberg et al. 2005), and higher socioeconomic status (Kringlen et al. 2001, Lundberg et al. 2005) have been associated with active participation in population-based studies in general. In the Health 2000 Survey, being married or cohabiting was associated with higher participation, but not the level of education (Koskinen et al. 2008), and in the FINRISK surveys lower socioeconomic groups were to some extent over-represented among non-participants compared with participants (Harald et al. 2007). Being married was the only socio-demographic factor affecting participation in original study II, with participation rates of married or divorced subjects being significantly higher than those of unmarried subjects.

Participation in a previously conducted epidemiologic study (31-year follow-up) was highly related to participation in the psychiatric field study conducted three years later, which agrees with the finding of earlier participation predicting later participation in longitudinal studies by e.g. Drivsholm et al. (2006).
Illness-related factors associated with participation

Schizophrenia has been associated with non-participation in population-based studies by e.g. Susser et al. (2000) and Lundberg et al. (2005). In original study II, non-participants were more often patients with schizophrenia than with other psychoses. Non-participants were shown more likely to have substance use disorders than participants according to Lundberg et al. (2005). In the original study II, though, non-participants had substance use disorders of the same magnitude as did participants.

Abnormalities in brain morphology in schizophrenia were studied in a large prospective study of first-episode schizophrenia known as the Hillside Hospital Prospective Study of First Episode of Schizophrenia (Lieberman et al. 1992). Out of 219 eligible subjects, 59 refused to participate and 68 were excluded for clinical or administrative reasons, leaving 92 subjects (58% out of 160 subjects) to enter the study. In original study II, 56% of the subjects with schizophrenia participated in the psychiatric field study. Altogether 48% of the patients with a baseline MRI scan in the Hillside Hospital Prospective Study of First Episode of Schizophrenia had a follow-up MRI scan after at least 12 months. According to Lieberman et al. (2001), those with only a baseline scan did not differ significantly from those with both a baseline and a follow-up scan in duration of psychotic symptoms prior to study entry, baseline CGI score, age at onset, or days to treatment response.

In a study of aggressive behaviour in schizophrenia, participants had a longer duration of illness, but they were less severely ill than non-participants (Cheung et al. 1997). In original study II, there was no difference in age at illness onset between participants and non-participants. The results from original study II are in accordance with those of Drivsholm et al. (2006) as far as non-participants having episodes of psychiatric hospitalisation more often than participants.

Non-participants have been shown to have more unhealthy behaviour, such as smoking (Drivsholm et al. 2006, Atherton et al. 2008). Also aspects of health and behaviour in youth have been shown to relate with non-participation (Pietilä et al. 1995). Such associations were not found in original study II, though. In the study of young men in the NFBC 1966 by Pietilä et al. (1995), the response rate was low among those who had mental disorders and who were unemployed or on disability pension.

It may be that subjects with psychiatric disorders do not want to participate owing to not wanting to be approached because of their illness, especially if they
are in remission. On the other hand, others may be pleased that someone is interested in how they are doing. Some patients may have had bad experiences with medical care in the past, e.g. involuntary care, or they may feel they have been treated badly and therefore do not want to participate in health studies. In longitudinal studies with multiple follow-ups, some may also have had bad experiences in the previous phases of the data collection. Recruitment is more effective in the ongoing ten-year follow-up of the psychiatric field study of the NFBC 1966. Subjects with a psychosis are interviewed in their home if they are not willing to participate otherwise. They are also brought to Oulu University Hospital by trained psychiatric nurses if they otherwise would not have the MRI scan taken.

6.2.3 Original study III

Only being on a disability pension due to a psychosis affected participation to a statistically significant extent, whereas participants tended to have received antipsychotic medication more often, had less treatment days due to psychosis and had a later onset of their illness. This supports the assumption that illness-related auxiliary variables may affect participation.

Sufficient evidence for brain volumes being different in subjects with a more severe course of illness than in those with a less severe course was not found in original study III. Mean GM volumes were smallest and mean CSF volumes largest in subjects receiving a disability pension, who did not have any antipsychotic medication, who had the most treatment days due to a psychosis, or whose age at the onset of the psychosis had been under 25 years. These results are in concordance with the studies by Rossi et al. (2000), Staal et al. (2001) and Milev et al. (2003), according to whom poor-outcome patients have more structural brain abnormalities than good-outcome patients.

Bias caused by differential participation may affect the estimates of brain volumes. It was assumed in original study III that the more severe the illness is, the greater the GM and WM volume loss. Based on this assumption, the results support the hypothesis that the use of IPW with illness-related variables may improve the brain volume estimates, as it reduced the estimated mean volumes of GM and WM in the schizophrenia group and increased the estimated mean volume of CSF in subjects with schizophrenia. The methodological issues are discussed later in this thesis.
6.2.4 Original study IV

In original study IV, the precision of self-reported data on medication use was substantial for antipsychotics and antidepressants, almost perfect for antidiabetics and antiepileptics, and moderate for beta-blocking agents. These results agree with previous studies. Agreement between self-reported and pharmacy data has been shown to be good for chronically used drugs or drugs used for serious conditions (Monster et al. 2002, Caskie et al. 2006). Agreement between personal interview and register-based data has been shown by Haukka et al. (2007) to be good for most psychotropic drugs, and substantial for antipsychotic and antidepressant medications according to Nielsen et al. (2008). Agreement between medication adherence in self-report and pharmacy records seems to be age-dependent, e.g. according to Guénette et al. (2005), interviews of elderly people exhibited poor agreement with pharmacy records.

When evaluating psychoactive medication use, people may be reluctant to report its use (Nielsen et al. 2008) or the psychological indication itself may lead to poor recall (Cotterchio et al. 1999). However, in original study IV the precision of self-reported data on psychoactive medication use was substantial. It may be, as Peters et al. (1998) suggested, that sensitive questions do not necessarily influence the response rate of the survey or the accuracy of the data.

Data on the use of certain medications can be accurately collected using either self-reported questionnaire data or pharmacy registers (e.g. antidiabetics). The results of original study IV suggest, however, that collecting data on the use of some other medications (e.g. beta-blocking agents) by using only pharmacy registers may lead to underestimation of the prevalence of medication use. When possible, it would be useful to gather information from various sources, in order to capture all the medications used.

Effect of gender, education and marital status on the reliability of self-reported data

In studies by van den Brandt et al. (1991) and West et al. (1995), gender was not shown to influence recall of drug consumption, which is concordant with the results on psychoactive and antidiabetic medication self-reports in original study IV. However, men reported the use of beta-blocking agents better than women.

Being married (Cotterchio et al. 1999, Caskie & Willis 2004) and having a higher level of education (Klungel et al. 2000) have been associated with better
self-reporting. In original study IV, married subjects reported the use of antipsychotics and antidepressants more precisely than those who were not married, and subjects with a secondary or tertiary education reported the use of antidepressants and antiepileptics more precisely than those with a basic education. Marital status or level of education did not affect reporting of antidiabetics and beta-blocking agents.

Effect of non-participation on the reliability of self-reported data

According to Drivsholm et al. (2006), subjects who participate in health studies have better health profiles than those who do not participate. In original study IV, a significant difference was found between participants and non-participants in the proportions using antipsychotic and antidepressant medications. Hence, non-response bias may affect the quality of data as well as recall bias. It may be that self-reported data underestimates the prevalence of the use of certain medications, and using register data, if possible, would increase the reliability of the research results.

6.3 Methodological considerations

6.3.1 Participation in the studies of the NFBC 1966

It is important to make an effort to ensure minimization of non-participation during the design phase of a study. In principle, there are several methods available for attempting to at least partially correct biases caused by selective non-participation. In a simulation study evaluating the validity of four methods for handling the problem of non-participation (complete-case analysis, weighting, regression imputation, and MI), Kristman et al. (2005) found that imputation improved estimates at levels of 25–40% attrition. However, they pointed out that none of the methods they tested would necessarily be useful in correcting for non-response bias in cohort studies.

In the National Survey of Health and Development 1946 (NSHD), the lowest rates of successful contact were in the early adult years (Wadsworth et al. 2003). Wadsworth et al. (2003) suspected that it could be due to loss of contact with cohort members because of changes of addresses and names. It may be, though, that it is easier to collect information on people in Finland than in many other
countries due to personal identification codes and a large amount of register data. The study members of the NSHD also had a chance to choose whether to respond for the first time in their early adult years. In the NFBC 1966, written consent was asked for as a part of the 31-year follow-up study in 1997, and the goals of the study were briefly introduced. This could have raised interest in participation. The sample of original study II was collected soon after, during 1999–2000.

There are ways to increase and encourage participation in research, e.g. monetary incentives (Edwards et al. 2002). As stated earlier, subjects participating in the 31-year follow-up were not paid, but in the psychiatric field study, the travel expenses of the subjects were paid and they received a daily allowance and a snack on the examination day. In Chapter 5, Section 21, of the Medical Research Act (Medical Research Act 1999) by the Ministry of Social Affairs and Health, the following is stated concerning remuneration of research subjects: No payment for participating in the research shall be made to the research subjects, their guardians, close relatives, any other persons closely connected with them, or to their legal representative. However, an appropriate remuneration may be paid to cover expenses or loss of earnings or for any other inconvenience suffered as a result of the research.

As the main focus in the psychiatric field study was on subjects with a psychosis, much effort was made to recruit them. They were sent more reminders than control subjects, and they were given more flexibility in e.g. changing the appointment than control subjects. If a subject with a psychosis could not complete all parts of the study during one day, he or she was given a chance to continue the study another day. None of the interviews were done at the subjects’ homes, but at least one patient was brought to the study centre. Altogether, this led to the participation rate of subjects with a psychosis being as high as that of non-psychotic control subjects.

6.3.2 Statistical methods for dealing with non-participation

The effect of non-participation is rarely corrected in psychiatric studies, and one should be cautious when interpreting non-corrected estimates. Using complete-case analysis entails substantial disadvantages. In addition to reducing the sample size, it is likely to bias the estimates. Complete-case analysis is justifiable only if the proportion of missing cases is low. The degree of bias depends not only on the proportion of missing cases, but also on the differences between participants and non-participants, and on the parameters of interest (Little & Rubin 1987).
Mazumdar et al. (2007) state that even though methodological procedures for avoiding bias caused by missing data are available, there is a gap between the theoretical developments and their practical implementations. Original study III aimed to give one practical example based on previous work with the same sample by Tanskanen et al. (2009).

According to Newgard & Haukoos (2007), multiple imputation is computationally straightforward, versatile, relatively easy to apply, and increasingly available in standard statistical software applications. However, Höfler et al. (2005) stated that weighting is the most appropriate approach for scenarios where entire data sets are missing, except for e.g. some basic socio-demographic information. They presented ways to use weighted data as a means of dealing with non-participation, with a presentation of statistical software available for that purpose.

**Use of the weighting method**

As shown in original study III, IPW is a reasonably easy way of overcoming the problem of missing data in the case of unit non-response. To use any weighting method, comparable data on all the eligible subjects intended to participate in the study are required. Cohort studies and clinical trials probably include a considerable amount of information collected at earlier phases of the study, or it may be possible to use register data. If not, general population data, e.g. census data, can be used to calculate weights. It should be noted, however, that the quality of weighting depends on the quality of auxiliary variables and models built on them.

Schizophrenia research is often based on clinical settings, where subjects with schizophrenia likely have a more severe course of illness than those living in the community. Population-based studies are important in order to achieve a more extensive understanding of schizophrenia, e.g. with regard to the natural course of the illness and its outcome. In population-based samples subjects may have widely differing diseases in terms of severity or the course of the illness. In view of this, estimated brain volumes in subjects with schizophrenia may differ between population-based and clinical samples. If the most severe cases drop out of population-based studies, statistical differences between patient and comparison groups may be weakened.

In original study III we presented an example of using IPW to adjust for non-participation in an MRI study on subjects with schizophrenia. The presented
results could be useful in longitudinal population-based or other follow-up studies, which contain data collected in the previous phases of the study. IPW could also be useful in cross-sectional studies if register or census data are available and in randomised clinical trials where drop-out rates are often high. Martin et al. (2006) found a median drop-out rate of 40% in their systematic review of antipsychotic medication trials, for example.

6.3.3 Reliability of self-reported data

The type of medication may affect the reliability of self-reported data. For instance, in original study IV, use of antidiabetics was reported more precisely than use of other medications, and use of beta-blocking agents less precisely. This may be due to antidiabetics being used daily on a long-term basis. Beta-blocking agents, on the other hand, are used for asymptomatic chronic illness, and may be used occasionally, on a per-need basis. It may also be that beta-blocking agents are too cheap to be entitled to compensation and hence do not appear in the pharmacy data.

The structure of the questionnaire may lead to some discordance between self-reported and pharmacy data. When planning a study one should be explicit in inquiring for data. In original study IV, data were gathered from a postal questionnaire in which subjects were asked about their current medication use. For some people current medication use may strictly mean regular use of a drug at a specified time point, whilst for others it may mean having a prescription to be used if needed.

According to Klungel et al. (1999), the accuracy of data was better for questions about drugs used for a specific indication than for open-ended questions. It would be useful to use e.g. the ATC to classify drugs already when preparing the questionnaire. If a subject recalls the use of a specified class of drugs, the name and dose of the drug and the duration of the drug use can be inquired in more detail. In original study IV, the subjects were asked to name the medications they used currently, but the names of possible drugs were not given. According to Cotterchio et al. (1999), listing the most commonly used drugs for specific indications could improve recall in questionnaires, and according to Boudreau et al. (2004), showing photographs of commonly used drugs or using calendar of life events could improve recall in interviews.

‘Current medication use’ was asked, hence no time-based recall bias should exist (West et al. 1995). In the analyses a modified 6-month fixed-time method
was used, which according to Nielsen et al. (2008), could be expected to capture well both chronically used and in-need medications.

Even though subjects reported the use of psychoactive medication accurately, some of them may have been reluctant to report their use. People may either be reluctant to report the use of psychoactive drugs, or the psychological indication itself may lead to poor recall (van den Brandt et al. 1991, Cotterchio et al. 1999). Nielsen et al. (2008) suspected possible ‘self-stigmatisation’ in self-reporting of antipsychotic medication, but they found no indication of such. Guénette et al. (2005) emphasised the confidentiality of collecting self-reported data in order to avoid over-reporting due to ‘social desirability bias’. However, the confidentiality of self-reported data also diminishes under-reporting due to self-stigmatisation.

6.4 Strengths and limitations

6.4.1 Original studies I-II

The NFBC 1966 is a fairly large population-based study with the opportunity to utilise data from nationwide registers, such as those on hospitalisations, education, pensions and medication use. We have the same register information on non-participants as we have on participants. The data of the NFBC 1966 have been collected since the birth of the cohort members (Rantakallio 1969), and the study deals with the personal health of the subjects. Perhaps because of this, non-participation was relatively minor in original studies I and II.

The data on psychiatric disorders were collected from the FHDR, which has been shown to be a reliable source of data on psychoses (Moilanen et al. 2003, Arajärvi et al. 2005). There is an implication, though, that subjects with psychiatric disorders were, on average, severely ill and thus the results in original study I do not represent the whole spectrum of mental health problems. There may not be information on e.g. all substance use or mood disorders, as they seldom lead to hospitalisation, despite being quite prevalent in the population (Lehtinen & Joukamaa 1994, Sobell et al. 1996). Furthermore, if a patient has a severe psychiatric disorder, e.g. schizophrenia, co-morbid disorders, such as depression, are not always diagnosed.

The general population-based samples in original studies I and II may have attenuated the differences between participants and non-participants. Reasonably small sample sizes also attenuate the power of the studies. It may be that some of
the statistically significant differences between non-participants and participants in original study II may be due to several comparisons. The comparisons were univariate. However, the factors affecting participation inter-correlate, so multivariate analyses could have given different results.

The prevalence of any hospital-treated psychiatric disorder in the whole cohort was 4%. Hence the effect of non-participation due to psychiatric disorders of this kind on many interesting health outcomes in the whole cohort is of minor significance. In studies specifically targeting persons with mental health problems, however, this is a major issue demanding adequate approaches to dealing with the induced missing data problem when analysing the data.

The 31-year follow-up was conducted in a particular population limited by region, time, and age, using specific arrangements and procedures for data collection. In spite of this, it is not unreasonable to believe, with due reservations, that the fact that the participation rate among those with a history of psychiatric disorder was lower than among those without one, would also hold in other analogous settings with a sufficiently similar socio-cultural and healthcare context and study setup. However, one should be very cautious in any quantitative extrapolation of the observed differences in the participation rate.

### 6.4.2 Original study III

As pointed out in a review article by Harris et al. (2009), editors and reviewers of psychiatry journals require that missing data should be treated properly in the analyses, using a statistically appropriate method. Original study III gives an example of how to use IPW in a psychiatric field study with missing data.

IPW can be used in many study settings, and can be applied to subgroups, as was done in original study III. Weights can also be calculated for both cases and controls, e.g. concerning education, and multiple auxiliary variables can be used simultaneously for calculating weights, e.g. on education and gender. When using IPW, the model used for response propensity should be built up carefully in order to avoid extreme propensity values, which would create highly variable weights and thereby highly variable weight-adjusted estimates (Potter et al. 2006). Auxiliary variables should be chosen in a way that leads to the most adequate weights. The measures of the severity of illness used in original study III were not necessarily very sensitive, whereas baseline symptom scales, for example, could be more sensitive measures.
The disadvantage of IPW, or other weighting methods, is that although they may reduce bias, the precision of the estimates may also be reduced due to increasing variance (Little et al. 1997). Little et al. (1997) concluded, though, that bias, which is often left unmeasured, is a more severe problem than increased variance, and hence weighting is justified. Haukoos & Newgard (2007) emphasised that a weighting methodology works best on fairly large samples, where the potential for bias is of more concern than the loss of precision. In original study III variance increased at most by 6.4% in univariate and 7.7% in multivariate IPW. Even though the current sample was relatively small, IPW changed the estimates of brain volumes without greatly increasing the variances of the estimates.

6.4.3 Original study IV

In original study IV, extensive pharmacy data were used by linking them individually to self-reported data. However, there was no information on the dose or amount of medication purchased, so the duration of the prescription could not be estimated. Due to this, all the information from the questionnaire in which the subjects had also reported the dosage of the medications they reported using could not be used. Assumptions had to be made about subjects using medication for a certain time period and about their state of medication use at the time of filling in the questionnaire. In addition, the pharmacy data only give information about the purchase of medication, not whether the medication has actually been consumed.

Haukka et al. (2007) suggested in their population-based genetic study of schizophrenia that medication use is rarely over-reported. They obtained current medication use by interview and from the prescription reimbursement database. In original study IV some of the subjects reported using medication, but had no data on their purchase. Part of the over-reporting may be explained by medication not being entitled to compensation, part of it may be due to the time window of addressing self-reported data to pharmacy data. In addition, although unlikely in this study, there may be a small proportion of people who are misusing drugs or who are otherwise unwilling to register their drug use. They are hardly likely to report their use in questionnaires, either. However, considering the large number of subjects studied in the cohort, over-reporting was minimal.

Even though the primary interest of original study IV was in psychoactive medication, certain drugs could be used for various disorders. Antiepileptics, for example, are mainly used to treat epilepsy, but they are also used as mood
stabilisers for treating bipolar disorders. New antidepressants may also be used for other disorders besides depression.

The questionnaire did not show the date when it was filled in. The date of either informed consent or the clinical examination (i.e. the second part of the 31-year follow-up) was used instead. Subjects may have filled in the questionnaire long before returning the consent form or before attending the clinical examination, which is why the six-month gap between the dates of purchase and assumed filling in of the questionnaire was chosen. This may have caused discordance between the self-reported and pharmacy data, but it also caused some uncertainty about the true state of medication use.
7 Conclusions

7.1 Main conclusions

This study revealed some associations between participation and the generalisability of the research, and the reliability of self-reported data collected by postal questionnaire from a large population-based birth cohort. Researchers should make an effort to retain participation rates as high as possible through careful planning and recruitment of subjects they intend to sample. For several reasons it is hardly ever even possible to recruit all the subjects from the target population.

In original study I, subjects with hospital-treated psychiatric disorders participated less actively than subjects without such disorders. The participation rates decreased uniformly from the phase of a simple postal questionnaire study to the more demanding phase of additional psychometric assessments. Due to non-participation, the true prevalence of psychiatric disorders is probably higher than the prevalence estimates from epidemiological studies when based on the data provided by the participants only. As a result of original study I, the recruitment of subjects who are least likely to participate could be strengthened. To maximise participation and retain data representative of the target population, there could be an alternative study protocol for subjects with e.g. a psychiatric disorder or low education. To collect the most important information, those who do not reply after several reminders should be offered an abridged questionnaire or interview. Also, when possible, register data or data collected earlier could be used in order to allocate resources most effectively and to identify groups that may require more recruitment effort.

In original study II, non-participants were more often patients with schizophrenia than with other psychoses. It may confirm that some results from the NFBC 1966 may be underestimated due to non-participation. This bias may also be present in other population-based studies on psychoses. When the focus of the study is on groups of subjects that are unlikely to participate, it is easier to target more effort on recruiting them than in general epidemiologic studies. It may be difficult for e.g. someone with a psychosis to participate in the study if he or she has to travel using public transportation. In such a case the research personnel can pick up the patient from his or her home and bring him or her to the research centre, which has been done in the ten-year follow-up of the psychiatric field.
study. If the patient does not want to participate in the clinical examination, he or she can be interviewed at home in order to collect the most essential information.

When missing data occur, they should be treated properly using a statistically appropriate method. Original study III gave an example of one useful method, namely inverse probability weighting, that can be used to improve estimates affected by non-participation in order to reflect true differences in the target population. A disability pension and more psychiatric hospitalisation due to a psychosis were the most effective auxiliary variables when estimating brain volumes in original study III. Based on this study, IPW can be taken into consideration in future studies, as the method can be used to handle the problem of selected non-participation in the case of unit non-response. When interpreting the results of the research, the possible effect of missing data should be taken into consideration. Also, whether missing data have been treated adequately should be evaluated.

The results of original study IV show that data collected by postal questionnaire can be assumed accurate enough for study purposes, even though they may underestimate the prevalence of medication use due to non-participation. Using several sources of information should be considered when collecting data, as certain medications (beta-blocking agents) were not fully covered in the pharmacy data, but some other medications (antipsychotics and antidepressants) were more often purchased by the non-participants. Special attention should be paid to the structure and phrasing of questions during the design phase of the study in order to cover the area of interest explicitly.

### 7.2 Future research

During the work on this doctoral thesis some new questions on the methods used to adjust for non-participation were raised. In the future it is important to study the effects of non-participation more deeply and introduce methods for handling missing values in various study settings. Specifically, evaluating the use of weighting methods in simulation studies with complete data and several arbitrarily generalised incomplete data sets will be of interest. Comparing these data sets gives valuable information on the quality of the weighting method and the level of adjustment gained from using such a method. Use of multiple imputation in psychiatric research is also important in the future.

The NFBC 1966 offers various possibilities to study these methods. A large amount of data has been collected in the previous phases of the study. The ongoing
ten-year follow-up of the psychiatric field study was started in 2008, and a 45-year follow-up of the whole cohort has been planned. Recruitment has been improved in the ten-year follow-up of the psychiatric field study, and methods for maximising participation of subjects with a psychosis are being used. The participation of those who participated in the previous psychiatric field study is essential in the follow-up, as the aim of the study is to evaluate the course of a psychosis, e.g. changes in symptoms and brain morphology. For instance, it is possible to evaluate how the improved recruitment protocol affects participation rates and the representativity of the sample. Also, in the 45-year follow-up of the whole cohort, representativity can be improved by allocating more effort to recruiting subjects with e.g. low education or a psychiatric illness.
References


Williams B, Irvine L, McGinnis AR, McMurdo MET & Crombie IK (2007) When "no" might not quite mean "no": the importance of informed and meaningful non-consent: results from a survey of individuals refusing participation in a health-related research project. BMC Health Serv Res 7(59).
Original publications

This thesis is based on the following articles, which are referred to in the text by their Roman numerals.


The published original papers are reprinted by permission of SAGE Publications Ltd. (I), Springer Medizin (II), and Wiley-Blackwell (IV). Also unpublished data (III) is presented in this thesis.

Original publications are not included in the electronic version of the dissertation.
1032. Niinimäki, Maarit (2009) Medical compared with surgical management in induced abortions and miscarriages
1033. Yan, Ying (2009) The antichlamydial effects of drugs used in cardiovascular diseases
1034. Sipola, Annina (2009) Effects of vascular endothelial growth factor (VEGF-A) and endostatin on bone
1038. Seppinen, Lotta (2009) The roles of collagen XVIII and its endostatin domain in wound healing, hair follicle cycling and bone development
1041. Liimatta, Erkki (2009) Intracellular calcium, preconditioning and regulation of cellular respiration in heart

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/
Marianne Haapea

NON-RESPONSE AND INFORMATION BIAS IN POPULATION-BASED PSYCHIATRIC RESEARCH

THE NORTHERN FINLAND 1966 BIRTH COHORT STUDY