Case counting considered

Rijnders, C.A.Th.

Publication date:
2008

CASE COUNTING CONSIDERED

Psychiatric Epidemiology and Clinical Judgement

C.A.Th. Rijnders

CASE COUNTING CONSIDERED

Psychiatric Epidemiology and Clinical Judgement

C.A.Th. Rijnders

Acknowledgement
We wish to express our gratitude to the Prevention Foundation (grant 28-816-2) and the Ministry of Health, Welfare and Sport (grant VWS-64-4528-545272-299857.98), which provided the resources for collecting the data generated by the Nijmegen Health Area-2 project.
The Department of Social Medicine and Public Health of the Radboud University Nijmegen and the Institute for Mental Health Care Midden-Brabant (Tilburg) are greatfully acknowledged for their financial support.

Cover design: Helga Tripp, Cees Rijnders

Printed by:  Offsetdrukkerij Ridderprint B.V. , Ridderkerk

e-mail: info@ridderprint.nl
http//www.ridderprint.nl

ISBN/EAN: 978-90-5335-154-3

Case counting considered

Psychiatric epidemiology and clinical judgement

Proefschrift
ter verkrijging van de graad van doctor
aan de Universiteit van Tilburg,
op gezag van de rector magnificus,
prof.dr. F.A. van der Duyn Schouten,
in het openbaar te verdedigen
ten overstaan van
een door het college voor promoties
aangewezen commissie
in de aula van de Universiteit
op maandag 2 juni 2008 om 16.15 uur

door

Cornelis Aloysius Theodorus Rijnders
geboren op 21 februari 1956 te Tilburg

Uitgeverij Ridderprint

2008

Promotor : Prof.Dr. P.P.G. Hodiamont
Copromotor : Dr. J.W. Furer

Content

1 Introduction 11

Part I 15

2 Epidemiology 17
2.1 Definition and background 17
2.2 (Meta-)theoretical aspects 20
2.2.1 Causality: the ultimate goal 20
2.2.2 Concepts of causality 21
2.2.3 Levels of causation 23
2.3 Operational aspects 26
2.3.1 Choice of determinants 26
2.3.2 Designs 27

3 Diagnostics in psychiatry 33
3.1 Definition and background 33
3.1.1 The idiographic approach to patients 33
3.1.2 The nomothetic or classificatory approach to patients 35
3.2 Meta-theoretical aspects 37
3.3 Theoretical aspects 39
3.3.1 Categories versus dimensions 39
3.3.2 Classification versus typology 40
3.3.3 Classification – its recent history 41
3.3.4 Classification – advantages and disadvantages 42
3.3.5 The psychopathological spectrum 43
3.4 Operational aspects 46
3.4.1 Operational methods of diagnosis 46
3.4.2 Diagnostic instruments 47
3.5 Psychometric evaluation of diagnostic procedures 49
3.5.1 Assessment of validity 49
3.5.2 Assessment of reliability 51
3.5.3 Sensitivity to change 53
3.6 Conclusions 54

4 Research objectives 59
4.1 Introduction: case finding and case identification 59
4.2 Descriptive epidemiology, i.e. case finding: the prevalence and distribution of identified cases 60
4.3 Diagnostics: case identification 61
4.4 Research questions 62

Part II 65

5 Monitoring: Psychiatric disorders in a Dutch health area – a repeated cross-sectional survey with the PSE 67
5.1 Introduction 67
5.2 Methods 70
5.2.1 The sample 70
5.2.2 Procedure 71
5.2.3 The general interview 71
5.2.4 The measurement of psychiatric disorder 73
5.2.5 Analyses 74
5.3 Results 75
5.3.1 Sociodemographic shifts 75
5.3.2 GHQ data 75
5.3.3 Relationship between GHQ and the PSE 76
5.3.4 PSE data 77
5.3.5 The relation between PSE data and sociodemographic variables 79
5.4 Discussion 84
5.4.1 Study limitations 84
5.4.2 Prevalence rates, GHQ – PSE relationship 84
5.4.3 Prevalence rates, the PSE-caseness 85
5.4.4 Sociodemographic shift and distribution of psychiatric disorders 86
5.4.5 Influences of urbanisation 88
5.4.6 Cohort, period, and age effects 88
5.4.7 Conclusions 90

6 Mapping: The Nijmegen Health Area-2 project 97
6.1 Introduction 97
6.2 Methods 101
6.2.1 History of the NHA-2 101
6.2.2 Study design 101
6.2.3 Survey sample 102
6.2.4 Interviews 102
6.2.5 Analyses 103
6.3 Results 105
6.3.1 Response and representativeness 105
6.3.2 Psychiatric prevalence 106
6.3.3 Distribution of psychiatric disorders 108
6.3.4 Psychiatric comorbidity 113
6.3.5 The NHA-2 and the NEMESIS 114

6.4 Discussion 116
6.4.1 Study limitations 116
6.4.2 Psychiatric prevalence 116
6.4.3 Distribution of psychiatric disorders 117
6.4.4 Psychiatric comorbidity 117
6.4.5 Influence of the three newly determined categories 118
6.4.6 The NEMESIS 118
6.4.7 Conclusions 120

Part III 129

7 Psychometric properties of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN-2.1) 131
7.1 Introduction 131
7.2 Subjects and methods 133
7.2.1 Test-retest procedure 134
7.2.2 Standardised procedure 135
7.3 Results 136
7.3.1 Test-retest procedure 136
7.3.2 Standardised procedure 137
7.4 Discussion 141
7.4.1 Study limitations 141
7.4.2 Literature 142

8 The relevance of clinical judgement in the SCAN-2.1 145
8.1 Introduction 145
8.2 Methods 148
8.2.1 Design of the basic study 148
8.2.2 Operationalising clinical judgement in SCAN 148
8.2.3 Analyses 149
8.3 Results 151
8.3.1 The difference between results obtained with and without clinical judgement 151
8.3.2 The consequences of the omission of clinical judgement on epidemiological data 154
8.4 Discussion 162
8.4.1 Study limitations 162
8.4.2 The difference with and without clinical judgement 162
8.4.3 The consequences of omitting clinical judgement for epidemiological data 163
8.4.4 Conclusions 164

9 A comparison of the composite international diagnostic interview (CIDI) with the schedules for clinical assessment in neuropsychiatry (SCAN) in the general population 169
9.1 Introduction 169
9.2 Methods 172
9.2.1 Setting 172
9.2.2 Instruments and design 172
9.2.3 Participants 172
9.2.4 Statistics 173
9.3 Results 174
9.4 Discussion 180

Part IV 187

10 Research results and general discussion 189
10.1 Case definition, case identification and case finding in psychiatric epidemiology 189
10.2 Descriptive epidemiology, i.e. case finding: the prevalence and distribution of identified cases 192
10.2.1 Monitoring 192
10.2.2 Mapping 195
10.2.3 Conclusions regarding descriptive epidemiology 204
10.3 Diagnostics: case identification 206
10.3.1 Psychometric properties 206
10.3.2 Clinical judgement 207
10.3.3 Comparison of the CIDI and the SCAN 209
10.3.4 Conclusions regarding diagnostics 210
10.4 General conclusions and recommendations 212
10.4.1 General conclusions 212
10.4.2 Study limitations 212
10.4.3 Recommendations 213
10.4.4 Final remark 216

11 Appendix 227
11.1 Psychiatric disorders in a Dutch health area: a repeated cross-sectional survey 229
11.1.1 Introduction 230
11.1.2 Methods 231
11.1.3 Results 233
11.1.4 Discussion 238
11.2 Conversion table SCAN/PSE-10 to PSE-9 241

Summary 255

Samenvatting 261

Dankwoord 267

Curriculum Vitae 269

Notes 271

“Not everything that can be counted counts,
and not everything that counts can be counted.” – Albert Einstein

1 Introduction

History of the Nijmegen Health Area project

The roots of the Nijmegen Health Area project (NHA) go back to the end of the 1970s, when Dr. Ferd Sturmans, head of the Institute of Social Medicine of Nijmegen University, took the initiative of conducting a general population survey on the need for health care. As this need is determined by a multitude of factors, the project was organised on a multi-disciplinary basis, not just for the assessment of somatic, psychiatric and psychosocial problems, but also with an eye on the medical, sociological and psychological background of healthcare consumption. To account for the complex topics to be surveyed, a number of university departments were invited to participate. After a period of preparatory talks, the sociology and psychiatry departments of Nijmegen University agreed to take part. The survey focused on the distribution of somatic complaints, psychosocial problems, and psychiatric symptoms, and also on the resulting illness behaviour and the attitude of patients to their need for care. As far as the psychiatric part of the survey was concerned, the data collection was carried out in 1983 by means of the PSE-9. The Prevention Foundation provided the resources for the Nijmegen Health Area 1 (NHA-1) project. The results of this first NHA project, and especially the psychiatric data, came to the attention of the Ministry of Health, Welfare and Sport, and played an essential role in the New Report on Public Health, which was released in 1984.

In the early 1990s, Dr. Paul Hodiamont, who had been responsible for the psychiatric part of the NHA-1, was approached by a representative of the same ministry, who suggested a replication of the psychiatric part of the NHA-1 project by the original research team in order to generate evidence on secular trends in the distribution of psychiatric disorders. The Institute of Social Medicine was willing to provide the organisational framework, and the Institute of Psychiatry was equally willing on condition that a supplementary research theme concerning the chronic use of benzodiazepines was incorporated into the study design. In order to meet this requirement, the co-operation of regional general practitioners (GP) was essential. Consequently, the Department of General Practice of Nijmegen University was invited to participate in the study, and the respondents were recruited randomly from the GPs' patient registers. This recruitment method brought about two consequences, the first of which was advantageous and the second disadvantageous: the first was the possibility of creating a list of high-risk indicators for psychiatric disorders; the second was that the medical ethics committee judged that the epidemiological part of the study was no longer an open field survey in the general population, but a patient-oriented study. As a result, a detailed letter for obtaining informed

consent had to be drawn up, which had a discouraging effect on potential survey subjects.
The data for the Nijmegen Health Area-2 project (NHA-2) was collected from September 1997 to January 1998. The Prevention Foundation and the Ministry of Health, Welfare and Sport provided the resources for the data collection for the NHA-2 project.
At the same time, the Trimbos Institute (Netherlands Institute of Mental Health and Addiction) was given a government grant for the NEMESIS study, a longitudinal study on psychiatric morbidity, health service use, and need for care in the general population. An advisory committee chaired by Dr. Harry Rooijmans was formed to monitor the progress toward the separate goals of, and also the cooperation between, the two studies. The main differences between the NHA-2 and the NEMESIS were as follows: although both were epidemiological general population studies, the former was regional and cross-sectional, and repeated (after 14 years), and carried out with the Schedules for Clinical Assessment in Neuropsychiatry (SCAN), an observer-based interview, whereas the latter was national and longitudinal and was carried out with the Composite International Diagnostic Interview (CIDI), a respondent-based instrument. Furthermore, the NEMESIS, which is more or less a replication of the National Comorbidity Study carried out in the USA a few years ago, determined psychiatric morbidity over a more limited diagnostic range than that determined in the NHA-2.

Thesis

This thesis uses descriptive epidemiological data to focus on some aspects of the reliability and validity of psychiatric diagnoses, with the objective of improving the quality of the diagnostic process, and through this, to enhance the quality of future epidemiological information. It will do so by investigating the correlation between relevant variables and the results of an international standardised clinical psychiatric interview, in particular version 2.1 of the SCAN (SCAN-2.1). The data were collected as part of a repeated cross-sectional survey conducted in the Dutch Health Area of Nijmegen.

Four parts

After this introduction, in part one, we will elaborate on the theory and essential characteristics of epidemiology (see chapter 2) and on the principles of the diagnostic process and the psychometric parameters of various diagnostic procedures (see chapter 3). We will conclude part one in chapter 4 with the research questions addressed in this thesis.

In part two, which is on epidemiologic outcomes, we will present a model for mapping and monitoring epidemiologic data. First, in chapter 5 we will present an example of monitoring, since this was the initial motivation for performing the NHA-2 study. This monitoring example is illustrated by a comparison of the prevalence and distribution data generated by the PSE-9 in the NHA-1 study, with the similar data from the NHA-2 study. In chapter 6, the design, methods, and results of the NHA-2 study will be presented as an example of psychiatric mapping in which the contemporary clinical interview SCAN-2.1 was used as the interview instrument. The results from this study will be placed in both a national and an international perspective.

In part three on diagnostics in epidemiology, the diagnostic process used in the SCAN-2.1 will be evaluated from different points of view. In chapter 7 we will present data on the psychometric properties of the SCAN-2.1, the central clinical diagnostic instrument we used in the NHA-2 study. Subsequently, in chapter 8, we will address in greater depth the process of clinical decision-making implemented in the SCAN-2.1. Finally, in chapter 9, we will elaborate on the SCAN-2.1 data through a comparison of the two currently most frequently used psychiatric diagnostic instruments, the SCAN-2.1 and the CIDI-2.1.

In part four we will summarise the research results, follow these with a general discussion, and finally, set out the implications of our results for future research methods and goals.

Part I

2 Epidemiology

2.1 Definition and background

According to the Dictionary of Epidemiology (Last 1995), today epidemiology is defined as "the study of the distribution and determinants of health-related states or events in specified populations and the application of this study to control health problems". 'Study' encompasses surveillance, observation, hypothesis testing, analytic research, and experiments. 'Distribution' refers to analysis by time, place, and classes of affected people. 'Determinants' are all the physical, biological, social, cultural, and behavioural factors that influence health. 'Health-related states and events' include causes of death, diseases, behaviour, reactions to preventive regimens, and provision and use of health services. 'Specified populations' are those groups of people with identifiable characteristics. 'Application to control health problems' makes explicit the objectiveof epidemiology: to promote, protect, and restore health. The concepts, principles, research strategies, and methods of psychiatric epidemiology are those of general epidemiology, adapted for the study of psychiatric disturbances, manifest in both behaviour and mental life (Anthony et al. 1995).

Epidemiology offers some of the best available research strategies for addressing critical questions in different areas of interest. First, it serves (mental) health care, as it attempts to answer questions concerning the nature, aetiology, and course of (psychiatric) disorders in order to improve treatment potential, prevention and planning. Furthermore, epidemiological evidence is indispensable for ethical, well-planned, well-considered, and financially sound political decisions in a broad range of policy-making areas. These include the planning and financing of (mental) health care facilities, disaster management, and penitentiary services.

There have been many definitions of epidemiology (Fleming & Hsieh 2002; Hodiamont 1996; Kluiter & Ormel 1999). In the course of time the definition has been broadened from concern with communicable disease epidemics to take in all phenomena related to health in populations. The word epidemiology is derived from the Greek: 'έπί' ('epi'), 'δήμος' ('demos'), and 'λόγος' ('logos'), which mean 'upon', 'people', and 'knowledge', respectively. Thus epidemiology originally meant 'the knowledge of what comes upon the people'. The ancient Greeks believed that plagues and other disasters (epidemics) were visited upon them by the gods or by destiny. The history of epidemiology that followed this initial conviction has been marked by dramatic transitions in thinking in response to new public health challenges and/or scientific breakthroughs (Susser & Susser 1996a; Susser & Susser 1996b). In the early nineteenth century, the industrial revolution brought about the Sanitary Era, with its focus mainly on societal factors and the theory of "miasma", a kind of polluting vapour

that emerged from the build-up of putrefying waste. Notwithstanding its evident success, sanitary epidemiology provided insufficient explanations as to how societal factors led to disease in individuals. It was not until 1854 that John Snow, a London medical doctor, systematically searched for the cause of a cholera epidemic in London (Bynum 1994; Vinten-Johansen et al. 2003). He identified the cause of the epidemic (not of the disease), and this made him the founding father of modern epidemiology. Intrigued by the clinical aspects of the cholera cases, he left the hospital, systematically visited the inhabitants of the afflicted Soho area, gathered information on their daily lives (observations) and in doing so, discovered what he never would have been able to if he had stayed at his desk: the casualties were mainly dependent on the Broadway pump for their water supply. He also concluded from his observations that the disease barely affected the employees of the local brewery, who were supplied with free beer. The pump was closed down and the epidemic ceased - a first act of political decision-making based on epidemiological evidence from geographic mapping. Although the exact aetiology of the disease was still unknown, the reason for which it was so specifically localised was discovered, and effective measures could be taken to stop this sneaky killer in its tracks.

At about the same time, the epidemiological method was successfully applied in psychiatry as well. Valdemar Steenberg, a Danish doctor working in a venereal clinic in Copenhagen, visited his colleague at the St. Hans Psychiatric Institute in the Danish capital. During the visit he recognised several patients with dementia paralytica (DP), whom he had treated for syphilis years before (Strömgren 1970). The epidemiological link between syphilis and DP was obvious to him, and he wrote a thesis on the subject. Some years later, Dr. Steenberg changed posts and started working at a peripheral hospital, where the prevalence rate of DP was much lower than in psychiatric in-patients in Copenhagen, where it was 45%. Those of his patients who did have the disease had all spent at least some time in the capital, at that time a centre of culture, pleasure and, as it seemed, venereal disease. A few decades later, at the end of the 19th century, Mattauschek and Pilcz, two Austrian army doctors, elaborated on these results and conducted a longitudinal survey up to 1911 on officers who had been infected with syphilis between 1880 an 1890. They discovered that, although syphilis was more or less endemic in the Austrian army, only 5% of the carriers developed DP. On analysing the difference between the groups without and with DP, it turned out that the former had had malaria fever soon after their syphilitic infection, in contrast to the officers who developed DP. Furthermore, they observed that in many tropical countries syphilis and malaria were endemic, whereas DP was practically unknown. They concluded that malarial fever probably had a preventive effect on the development of DP. Eventually, this led to the malarial treatments for which Wagner von Jauregg, as the first researcher in a psychiatric area, was awarded

the Nobel Prize in Physiology or Medicine in 1927 “for his discovery of the therapeutic value of malaria inoculation in the treatment of dementia paralytica” (Nobelprize.org).

Both case histories illustrate the importance of mapping, of drawing a detailed picture of the circumstances of a disease, as a basic tool for the epidemiological analysis of health-related problems. Repeated mapping, or monitoring, is gaining interest in the (mental) health care. That is so because this technique reveals fluctuations in disease frequencies in the population, facilitating the search for underlying causes and providing the rationale for preventive actions and facility planning, among other measures.

2.2 (Meta-)theoretical aspects

2.2.1 Causality: the ultimate goal

Common to all sciences are four general goals: *understanding, prediction, control* and *systematisation*. Moreover, the goal of understanding has two components, description and explanation. To achieve understanding, two basic types of observation are carried out: on the one hand, naturalistic observation, to *describe* phenomena, and on the other, experimental observation, to *explain* phenomena. These are central to descriptive research and explanatory or experimental research (Gould 2002). In epidemiology, these research objectives are pursued by means of three research strategies, each with its specific leading question. Descriptively, the question is, "*what condition* occurs in whom, where and when?". Analytically, (explanationwise) the question is, "*why* does the condition occur?". Both research strategies are means to the end of achieving control over diseases and disorders. Experimentally, the question is "what can be done to *prevent* or *heal* the condition?". All three strategies are illustrated by the example of DP as described above.

The foundation upon which science seeks to achieve these goals is comprised of a number of assumptions, notably with respect to *reality* and *determinism*.

When it comes to reality, scientists assume that objects and events have an existence external to and independent of man. In psychiatric epidemiology, constructs, such as schizophrenia, are, for the sake of convenience, given the status of facts. The true status of such constructs is, however, a complicated and controversial issue in psychiatry that deserves a review in its own right. We shall discuss this topic in chapter 3.

Determinism refers to the belief that the occurrence of an event or a behaviour is determined by (or dependent on) antecedent events. Because a phenomenon could have many, or only obscure, antecedents we might not currently know every, or even any, of its determinants, but that does not preclude the eventual discovery of all the causal factors. Schizophrenia, again, is an example of a mental disorder for which the causes have remained obscure for a long time. Now, however, we are gradually learning more about its biological and environmental determinants. Unravelling causality is the nec plus ultra of (psychiatric) epidemiology, since it offers the best chance of treating or preventing disorders effectively.

Mapping provides essential information about the *who, where* and *when* of the disorder in question and is an indispensable first step in the search for causality (Susser et al. 2002).

2.2.2 Concepts of causality

Basically, in epidemiological research studies the how is studied of the association of variables with one another (Kluiter & Ormel 1999). Generally speaking, reported statistical associations, such as those disclosed in mapping studies, can be explained as either artificial, indirect, or causal (aetiological) (Lilienfeld & Lilienfeld 1992). Epidemiology primarily aims at the clarification of the nature, the direction, and the power of causal relations. Causal relations in general medicine, let alone in psychiatry, can not be proven, but at best only be made plausible on the basis of circumstantial evidence (Kluiter & Ormel 1999). A strict time sequence (Y follows X) is necessary but not sufficient to verify a causal relation. Reactive change (when X changes, Y changes too) is a powerful indication, but difficult to assess in psychiatry. Furthermore, in psychiatry disorders are seldom assumed to be initiated by a single cause, which makes the recognition of causality even more difficult. From this perspective, Rothman (1986) developed the notions of sufficient and component cause, whereby the latter is a necessary but not sufficient condition in the explanation of the occurrence of a disease, whereas a sufficient cause is a unique constellation of component causes that inevitably leads to the onset of a disease.

The underlying influences on the incidence or duration of diseases are called determinants. A determinant has such an effect because either it is a component cause (or *factor*) of the disease itself, or it has an influence on one or more component causes (in which case it is referred to as an *indicator*). In the example of DP, living in an urban environment is an indicator, whereas having unprotected sex with prostitutes is a factor for contracting syphilis. Some indicators, however, do not have fixed meanings under all circumstances; in fact the contribution of some indicators depends on how they are interpreted in a socio-cultural context. A social indicator, which is more widespread in the population, is often more widely accepted and results in a lower stress level than an indicator that occurs incidentally. Murphy et al. (2003), reporting on the Stirling County study, concluded that “it was only when smoking was becoming non-normative that a strong and positive association with depression appeared”. Similarly, there is quite a difference between being an atheist in a large city and being an atheist in a small religious village.

A second and more qualitative, aetiological approach to causal factors is based on a differentiation between the *predisposing*, *facilitating*, *precipitating*, and *perpetuating* qualities (Hodiamont 1999). These qualities are not mutually exclusive, so that factors can be attributed to more than one of them. Predisposing factors are factors that increase the probability of an individual’s contracting a certain disorder. These vulnerability factors are by definition component, but they are rarely sufficient. In Snow’s example of the Soho cholera epidemic, living in the densely populated Broadway area is a

predisposing factor. Facilitating factors augment or abate the risk of onset of diseases, but differ from the former in that they are variable in character, and intervention is possible. In the cholera story, the poor, often open, sewer system and lack of hygiene (failure to boil water before it is drunk) are facilitating factors. A precipitating factor is related to the actual onset of a disease. When this factor is more important than others for the actual disease, it may be a component factor. An illustration is the drinking of unboiled water from the Broadway pump. Finally, perpetuating factors maintain the expression of the disease, e.g. the passage of contaminated faeces through the open sewer system of the Soho area, which polluted the Broadway water supply. Perpetuating factors can have either facilitating or precipitating potential or neither of them, in that they do not have the quality needed to influence the actual onset of the disease.

Currently, the approach to causality is incorporated into the diathesis-stress model (Butcher et al. 2007; Ingram & Luxton 2005; Kluiter & Ormel 1999; Meehl 1962; Zubin & Spring 1977). In these models, a predisposition to developing a disorder is termed a diathesis. It can derive from biological, psychosocial, and/or sociocultural causal factors, and the different viewpoints tend to emphasise the importance of different kinds of diatheses. Many mental disorders are believed to develop as the result of some kind of stressor or precipitating event that acts on a person who has a diathesis or vulnerability for that disorder. Stress, the response of a person to demands he or she perceives as exceeding his or her personal resources, usually occurs when that person experiences chronic or episodic events that are undesirable. Factors contributing to the development of a diathesis are themselves sometimes highly potent stressors (Brown & Harris 1978). There are several different ways that researchers have proposed in which a diathesis and stress can combine to produce a disorder. In the so-called additive model, the diathesis and the stress summate. The greater the underlying vulnerability, the less stress is needed to trigger the behaviour/disorder. Conversely, when there is a smaller genetic or early-life contribution, greater life stress is required to produce the particular result. In what is called an interactive model, a certain level of diathesis must be present before stress can have an effect. Thus someone with no diathesis will never develop the disorder no matter how much stress he or she experiences, whereas someone with a diathesis will be increasingly likely to develop the disorder with increasing stress. In terms of Rothman's causality model, the predisposition to vulnerability comprises the compilation of component causes that increase the risk of occurrence of a psychiatric disorder, either permanently or over a prolonged period, without being sufficient to actually elicit the disorder. An extra component cause (the stressor) is needed to trigger the onset of the disease. In the additive model, the stressor itself can be a factor that is sufficient to cause a disease. Even in healthy persons, for example, the cholera bacillus

itself can provoke the disease if it is sufficiently concentrated. Since the late 1980s, attention has been focused on the concept of protective factors, which modify a person's response to environmental stressors, making it less likely that someone with a certain diathesis will experience the adverse consequences of stressors (Rutter 1985). Protective factors, however, do not necessarily arise from positive experiences. Sometimes exposure to stressful experiences that are dealt with successfully can promote a sense of self-esteem that serves as a protective factor. Thus, paradoxically, some stressors promote effective coping. This "inoculation" effect is more likely to occur with moderate than with mild stressors (Rutter 1987). Protective factors sometimes lead to resilience – the ability to adapt successfully to even very difficult circumstances.

2.2.3 Levels of causation

The "level of causation" concept facilitates the specification of causes at different levels of organisation. Shortly after the Second World War, Von Bertalanffy (a biologist) introduced his general systems theory (Romme et al. 1981). This framework, an organisation of hierarchically ordered systems, created the possibility of integrating information from many different scientific areas. In general, there are seven system levels which can be used as a reference for measurement: the societal level, the group/community level, the individual level, the organ level, the tissue level, the cellular level, and the genetic level.

In the nineteenth century, in the Sanitary Era, when a social transformation (illustrated by industrialisation, amongst others ways) was associated with an accumulation of decaying waste resulting in polluting vapours, the societal level prevailed with its "miasma" paradigm. At the turn of the nineteenth century, the new science of microbiology heralded the Infectious Disease Era, which replenished the social process of infectious disease transmission with explanations of diseases on a more individual level on the basis of the "germ" paradigm. In the 1950s, the next shift was stimulated by the rapidly changing health profile of the western world. Thanks to antibiotics, infections were cured on an individual level and diseases were brought under control on a societal level, so that non-infectious, chronic disorders such as cardiovascular diseases and cancer became more apparent. It was not until after the Second World War that it was verified that smoking was a risk factor for lung cancer, which announced the Chronic Disease Era with its "risk factor" paradigm, which in turn signified a return of the societal level.

Today we think scientifically not only about individual risk factors, but also about the influence of society and the group (like the family and other social subgroups with their role patterns), and of tissue, cell and gene. Although efforts are increasingly being made to unravel psychiatric disorders on the basis of

integrated disparate reference levels, including, in particular, genetic and tissue levels (Susser et al. 2002), for psychiatry the societal, community and individual level are still the most prevalent.

As this thesis concentrates on psychiatry in its present-day state, in which societal and individual causal levels feature prominently, a few words on these levels of measurement are in place. A study that is based on a single level of causation has shortcomings that can be overcome by research on other levels.
As far as the individual level is concerned, there are three important limitations (Susser et al. 2002). First, when a factor (like hygiene in developed countries) is generally present in the study population, its influence cannot be adequately studied because of its lack of variation. Data on the impact of this ubiquitous factor are speculative, although it may be involved, for example, in the predisposition to, or precipitation of, a disease, and its removal may have an enormous effect on risk modification. Second, quite often, the prevalence of a specific risk factor on the individual level is influenced by societal change. Consequently, an explanation on the societal level is indispensable to an understanding of many changes in disease prevalence. For example, the reason for the increase in the prevalence rate of allergic reactions of individuals, found in many studies, is still not traceable and might be linked to an increase in the standard of hygiene or of sterility (these being societal factors) in developed countries. Finally, as mentioned in the previous paragraph, the attribution, and consequently the significance, of a risk factor on the individual level can be influenced by the socio-cultural context of a subpopulation, making the societal level necessary for an adequate appreciation. In conclusion, to understand a sufficient cause, usually the individual approach per se is not satisfactory, as the hygiene hypothesis shows in the example of the growing allergic constitution. All in all, it may even be this common risk factor that carries the greatest implication for disease prevention, although its contribution is undetectable in a pure individual design.
The societal level mainly helps to clarify determinants (factors and indicators) that are invariant within a population (not visible at the individual level) and also determinants in contexts and relations surrounding individuals. The distinction between the individual and the societal level is primarily a matter of magnitude, and consequently arbitrary (Susser et al. 2002). By definition, a society may constitute any combination of individuals who are connected in some meaningful way, from families to nations, the Roman Catholic Church and even the common denominator of developed countries. Each community or society has its particular attributes, which shape the experiences of people living in it. Although social factors appear to be remote from the occurrence of specific diseases in individuals, they are of great significance as causal determinants. While (specific) risk factor investigation can provide a more proximal cause

mechanism, research into societal factors may be more likely to indicate distal causes, which may lead to effective intervention strategies. For example, Snow, in his search for the cause of cholera, was not aware of the proximal, biological cause, but found a more distal cause on a communal level, which led him to the simple but effective solution of closing down the Broadway water pump.

With current research, including increasingly complex statistical processing of a growing number of determinants, it is not possible to accurately measure all important influences at all causation levels and consequently a fully comprehensive understanding of disease causality is still far away. Even at the individual and the societal level, a fully comprehensive measurement of important factors is still overly ambitious. Nevertheless, the object of this thesis is to contribute to the cumulative evidence that clears a path toward insight into the causal process.

2.3 Operational aspects

2.3.1 Choice of determinants

Mapping is increasingly used to examine the causes of mental disorders (Susser et al. 2002). The result of mapping concerns the correlation of quantitative frequency data with qualitative distribution characteristics (Anderson 1999). In epidemiological terms, the questions to answer are: what are the prevalence rates (quantitative data) and what is their distribution (qualitative characteristics). The latter, or the determinants, can be factors that are directly involved in, and that influence, the quantitative phenomenon, or indicators that simply point toward an indistinct, indirect relation. In the DP example, living in an urban environment is an indicator, namely a higher risk of contracting syphilis, and syphilis itself is a factor in the development of DP.

Of the many *qualitative* determinants at hand for making the diathesis-stress model operational, a selection of sociodemographic characteristics, coping styles, life events and other stressors is commonly used (Aalto-Setala et al. 2002b; Brugha et al. 2004; Brugha et al. 2005; Fryers et al. 2005; Roca et al. 1999; Roca-Bennasar et al. 2001; Wang 2005). The choice of the most meaningful determinants is still subject to trial and error, as the following two examples illustrate: (1) By way of operationalising the vulnerability/protecting factors, Aalto et al. (2002b) used trait anxiety, defence styles, self-esteem, life events and somatic symptoms, and found trait anxiety an important risk factor for mental distress, while gender differences suggested differences in coping; (2) Several research groups (Brugha et al. 2004; Selten et al. 2005; Selten & Cantor-Graae 2005) found that ethnicity as a factor is strongly associated with a higher prevalence rate of psychosis. But Brugha et al. established that the corresponding and underlying risk factors were socio-economic status (as measured by unemployment, poverty and lower social class), and social support and roles (represented by lone-parent status, low perceived social support, and having a small primary social support group). It should be noted, however, that none of these can be said to be a sufficient cause; until the causal correlation is clarified, it is better to talk about determinants.

The abundant array of determinants (amongst them the ones used by Aalto-Setala and Brugha in the aforementioned examples), which are by no means exhaustive, can be classified as in Table 2.3.1, in which the diathesis–stress model is related to the causal level of explanation. The vulnerability and protective traits are relatively stable factors, whereas stressors are more changeable over time.

We chose our determinants with the objective of investigating whether the more or less consistent relationship that has been found between psychiatric disorder

and sociodemographic characteristics (Dohrenwend & Dohrenwend 1974; Surtees et al. 1983) is also true for the Nijmegen Health Area Project.

Table 2.3.1
An example of sorting determinants into the diathesis-stress model and the level of causation

Diathesis stress Causal level	Vulnerability / protective traits	Stressors
Individual	(low) trait anxiety	life events
	coping styles	somatic symptoms
	self-esteem	state anxiety
	marital status	changing social roles
	social embedding	lone parenthood
	socio-economic status	loneliness / social deprivation
	education	advanced age
	gender	declining vitality
	chronic (somatic) disease	drug prescription
Societal/ communal	ethnic composition	diminished social support
	primary groups	poverty
	social class	unemployment
	economic growth/prosperity	multiple roles
	social cohesion	(imminent) natural disasters
	level of local services (water, health care, etc.)	population density (urbanisation)
	religious belief	

2.3.2 Designs

The history of psychiatric epidemiology, especially its designs, has been described in terms of generations (Anthony et al. 1995; Dohrenwend 1990; Dohrenwend & Dohrenwend 1982). Crucial to the typology of psychiatric epidemiological research designs are (1) the specificity of the psychopathological measurement used, (2) the quality of the sample, (3) the level of structure of the measurement, and (4) whether the score is observer- (objective) or respondent- (subjective) based. In other words, the research questions are: *what* can be said about *whom?,* what is the *reliability* of the information?, and what is its *clinical validity*?

With the first generation designs, basically retrospective, administrative, clinical treatment statistics were used to study the association of sociodemographic

variables with (more or less) specific mental disorders. Hence, the sample was restricted to persons referred for treatment (2), and the assessment of psychopathology, based on the disease model[i] used by the clinician (1 and 4) in a free-ranging diagnostic conversation (3), was not carried out with a reliable and standardised instrument. As a result, the data could not be generalised to the open population and were of limited value for policy decisions. The second generation designs sought to estimate the prevalence rate of general distress (1) in a sample of the open population (2), typically without much attention to specific psychiatric disorders (1), partially because a clear and reliable diagnostic categorical system was lacking. In that era, the data concerned the population as a whole, providing a basis for conclusions supporting policy decisions. But the outcome measure of general distress, mostly found by means of questionnaires (3), was based on the illness model[i] of (subjectively) perceived morbidity (4), which brought about a very high prevalence rate of general distress. From the viewpoint of clinical relevance (as part of the disease model), the prevalence figures seemed to be excessively high. The third generation design combined the field survey approach (2) with standardised diagnostic instruments (3) and the intentional focus on specific disorders (1). In respect of the source of the information (4), two traditions co-exist, the clinical/objective SCAN/PSE tradition and the respondent-based/subjective CIDI-DIS tradition. In short, the development of psychiatric epidemiology is characterised by a generalisation from the population sample in the denominator and also by a far-reaching specification of the disorders in the numerator of the epidemiological fraction. So that policy makers are provided with accurate data, the contemporary objective is to search for standardised instruments that are able to measure psychiatric disorders as reliably and categorically as possible. This process of specifying mental disorders will be discussed in the next chapter.

References

Aalto-Setala T, Poikolainen K, Tuulio-Henriksson A, Marttunen M and Lonnqvist J. (2002) Predictors of mental distress in early adulthood: a five-year follow-up of 709 high-school students. *Nordic journal of psychiatry,* 56, 121-125.

Anderson AJ. (1999) Methodological approaches in mental health services research and program evaluation. *International Journal of Psychosocial Rehabilitation,* 4, 73-92.

Anthony JC, Eaton WW and Henderson AS. (1995) Looking to the future in psychiatric epidemiology. *Epidemiologic Reviews,* 17, 240-242.

Brown GW and Harris T. (1978) Social origins of depression: a reply. *Psychological Medicine,* 8, 577-588.

Brugha T, Jenkins R, Bebbington P, Meltzer H, Lewis G and Farrell M. (2004) Risk factors and the prevalence of neurosis and psychosis in ethnic groups in Great Britain. *Social Psychiatry and Psychiatric Epidemiology,* 39, 939-946.

Brugha TS, Weich S, Singleton N, Lewis G, Bebbington PE, Jenkins R and Meltzer H. (2005) Primary group size, social support, gender and future mental health status in a prospective study of people living in private households throughout Great Britain. *Psychological Medicine,* 35, 705-714.

Butcher JN, Mineka S and Hooly JM. (2007) *Abnormal psychology*. (12 ed.) Allyn & Bacon, Boston.

Bynum WF. (1994) *Science and the practice of medicine in the nineteenth century*. Cambridge University Press, Cambridge.

Dohrenwend BP. (1990) 'The problem of validity in field studies of psychological disorders' revisited. *Psychological Medicine,* 20, 195-208.

Dohrenwend BP and Dohrenwend BS. (1974) Social and cultural influences on psychopathology. *Annual Review of Psychology,* 25, 417-452.

Dohrenwend BP and Dohrenwend BS. (1982) Perspectives on the past and future of psychiatric epidemiology. *American Journal of Public Health,* 72, 1271-1279.

Fleming JA and Hsieh CC. (2002) Introduction to epidemiologic research methods. In MT Tsuang en M Tohen (Eds.), *Textbook in psychiatric epidemiology, second edition*. pp. 3-33. Wiley-Liss, New York.

Fryers T, Melzer D, Jenkins R and Brugha T. (2005) The distribution of the common mental disorders: social inequalities in Europe. *Clinical practice and epidemiology in mental health,* 1, 1-12.

Gould JE. (2002) *Concise handbook of experimental methods for the behavioral and biological sciences*. CRC Press, Boca Raton, Florida, USA.

Hodiamont PPG. (1996) *Measuring with more measures; an exploration on concepts of social psychiatry [in Dutch]*. Tilburg University press, Tilburg.

Hodiamont PPG. (1999) Comorbidity [in Dutch]. In A de Jong, W Van den Brink, J Ormel en D Wiersma (Eds.), *Handbook on psychiatric epidemiology*. pp. 84-110. Elsevier / De Tijdstroom, Maarssen.

Ingram RE and Luxton D. (2005) Vulnerability-stress models. In BL Hankin en JRZ Abela (Eds.), *Development of psychopathology: A vulnerability-stress perspective*. Sage, New York.

Kluiter H and Ormel J. (1999) Psychiatric epidemiology: background, methods and techniques [in Dutch]. In A de Jong, W Van den Brink, J Ormel en D Wiersma (Eds.), *Handbook on psychiatric epidemiology*. pp. 19-69. Elsevier / De Tijdstroom, Maarssen.

Last JM. (1995) *A Dictionary of Epidemiology*. (third edition ed.) Oxford University Press, New York.

Lilienfeld AM and Lilienfeld DE. (1992) *Foundations of epidemiology*. Oxford University press, New York.

Meehl PE. (1962) Schizotaxia, schizotypy, and schizophrenia. *American Psychologist,* 17, 827-838.

Murphy JM, Horton NJ, Monson RR, Laird NM, Sobol AM and Leighton AH. (2003) Cigarette smoking in relation to depression: historical trends from the Stirling County Study. *American Journal of Psychiatry,* 160, 1663-1669.

Roca M, Gili M, Ferrer V, Bernardo M, Montano JJ, Salva JJ, Flores I and Leal S. (1999) Mental disorders on the island of Formentera: prevalence in general population using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Social Psychiatry and Psychiatric Epidemiology,* 34, 410-415.

Roca-Bennasar M, Gili-Planas M, Ferrer-Perez V and Bernardo-Arroyo M. (2001) Mental disorders and medical conditions. A community study in a small island in Spain. *Journal of Psychosomatic Research,* 50, 39-44.

Romme MAJ, Kraan H and Rotteveel R. (1981) *What is Social Psychiatry, an introduction [in Dutch]*. Samson, Alphen aan den Rijn.

Rothman KJ. (1986) *Modern epidemiology*. Little, Brown and Company, Boston.

Rutter M. (1985) Resilience in the face of adversity. Protective factors and resistance to psychiatric disorder. *British Journal of Psychiatry,* 147, 598-611.

Rutter M. (1987) Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry,* 57, 316-331.

Selten JP and Cantor-Graae E. (2005) Social defeat: risk factor for schizophrenia? *British Journal of Psychiatry,* 187, 101-102.

Selten JP, Zeyl C, Dwarkasing R, Lumsden V, Kahn RS and Van Harten PN. (2005) First-contact incidence of schizophrenia in Surinam. *British Journal of Psychiatry,* 186, 74-75.

Strömgren E. (1970) Anniversary lecture [in German]. In *150 Years 'Landeskrankenhaus Schleswig'*. pp. 20-27. Landeskrankenhaus Schleswig, Schleswig.

Surtees PG, Dean C, Ingham JG, Kreitman NB, Miller PM and Sashidharan SP. (1983) Psychiatric disorder in women from an Edinburgh community: associations with demographic factors. *British Journal of Psychiatry,* 142, 238-246.

Susser E, Bresnahan M and Link B. (2002) Peering into the future of psychiatric epidemiology. In MT Tsuang en M Tohen (Eds.), *Textbook in psychiatric epidemiology, second edition*. pp. 195-211. Wiley-Liss, New York.

Susser M and Susser E. (1996a) Choosing a future for epidemiology: I. Eras and paradigms. *American Journal of Public Health,* 86, 668-673.

Susser M and Susser E. (1996b) Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. *American Journal of Public Health,* 86, 674-677.

Vinten-Johansen P, Brody H, Paneth N, Rachman S, Rip M and Zuck D. (2003) *Cholera, Chloroform, and the Science of Medicine; A Life of John Snow*. Oxford University Press, Oxford.

Wang J. (2005) Work stress as a risk factor for major depressive episode(s). *Psychological Medicine,* 35, 865-871.

Zubin J and Spring B. (1977) Vulnerability: a new view of schizophrenia. *Journal of Abnormal Psychology,* 86, 103-126.

3 Diagnostics in psychiatry

3.1 Definition and background

Diagnosis is conceptualised as the identification of a disease, disorder, syndrome, condition, etc. Etymologically the term diagnosis is derived from the Greek word διάγιγνώσκω, which literally means to know through. This meaning has evaluated into two distinguishable connotations. First, when a diagnostician "looks through" a patient, he or she knows and understands all signs and symptoms in a persons individual context, leading to potential treatment strategies. Second, when a phenomenon is known through, two aspects of it are understood: the fact that a combination of certain signs and symptoms form a syndrome, and that this syndrome can be distinguished from other syndromes, which makes possible categorisation. This duality illustrates the fundamental polarity that bedevils all diagnostic systems, namely the polarity between the idiographic and the nomothetic point of view, given that an individual can be studied from each of them. In the context of the former, a patient is regarded as a unique subject, and any evaluation that ignores uniqueness will deplete the information required for individualised treatment. The latter refers to the way in which a subject is considered to resemble other people and to be governed by universal laws. In practice, no clinician can function solely at either pole. Although the assessment may be totally individualised, every patient has some characteristics in common with others. On the other hand, a comprehensive treatment plan demands more than a mere description for a stereotypic tag. Whereas idiographic diagnosis is highly useful for treatment planning, nomothetic or categorical diagnosis is highly useful for summary communication and prediction. It follows that the two approaches complement one another (Nurcombe & Gallagher 1986).

3.1.1 The idiographic approach to patients

In principle, medical doctors busy themselves with the treatment of individual ill persons and seek to explore their specific illness in depth. They seldom ask themselves what illness really is (Semple et al. 2005). For psychiatrists this question is more germane, for several reasons. More than with other illnesses, the diagnosis of psychiatric disorders is dependent on subjective experiences rather than on objective abnormalities, and in the diagnostic process, value judgements play an important role. Furthermore, psychiatric disorders may have social and legal implications that add to the burden of morbidity.

In the English literature, a meaningful distinction is made between illness, disease and sickness, with reference to subjective experience, the professional point of view and treatment, and social and legal implications, respectively.

Patients generally present complaining of or suffering from signs or symptoms (*illness*), a phenomenon called illness behaviour, related to the lay experience of being ill. Disease encompasses either a specific tissue lesion or a characteristic constellation of signs and symptoms, as diagnosed on the basis of clinical (professional) judgement, whereas sickness has to do with the social deficit consequent on symptoms. Each of these entities can exist without the other two, as shown in Figure 3.1.1 (Boot & Knapen 2001).

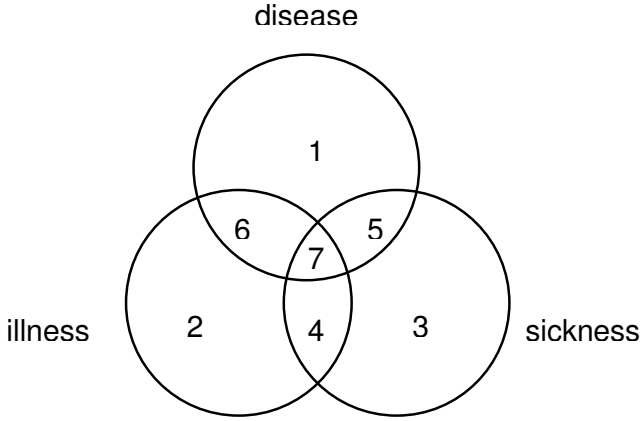

Figure 3.1.1

A patient with an asymptomatic form of cancer has a disease, does not feel ill, nor is he sick (see area [1] in Figure 3.1.1). Patients can experience symptoms, but may act as healthy persons, for example by not consulting a doctor [2]. Some patients may behave as if they were sick, for instance by simulating a disease without having an objective disorder or a subjective feeling of illness, and claim an entitlement pursuant to the Sickness Benefits Act [3]. Someone with chronic fatigue syndrome may experience himself as being ill (subjective) and try to qualify for a sickness benefit. This subjective experience, however, may not be acknowledged by the medical profession as (part of) an established syndrome, which is a formal criterion for granting a sickness benefit [4]. In [5] the patient has a specified disorder, behaves like someone with that disorder (for example, receives wages from her employer in accordance with the Sickness Benefits Act), but does not feel ill. Some patients feel ill and really have a specified disorder, but tend to act as if nothing were wrong [6]. Finally, [7] is the model patient, who experiences the illness (subjective), has a lesion as determined by medical criteria (objective), and behaves accordingly (sick role).

It should be pointed out in passing that individuals need not be sick in order to show illness behaviour, and even illness behaviour alone can be pathological. Illness and illness behaviour are subject to certain social conventions. An

attending doctor has the authority to grant a patient a sick role, and this role brings about two rights (a) and two obligations (b): a1) the right of exemption for the patient from his or her normal duties and a2) the right that the sick person is not held responsible for his or her sickness; b1) the obligation for the patient to seek medical professional help and b2) the obligation to comply with treatment in order to get well.
Consequently, not only do doctors treat signs and symptoms; they also legitimise the social implications of disease. It follows that they serve a dual function: that of medical practitioners and that of social gatekeepers.
To sum up, the psychiatrist, like other doctors, by way of achieving the goals of communication, prediction and control of treatment planning, has to fulfil the interactive roles of medical practitioner and social agent, with diagnosis as his or her indispensable tool.

3.1.2 The nomothetic or classificatory approach to patients

The act of classifying can be looked upon as a cognitive coping strategy: it enhances the chances of survival by reducing the chaos of incoming information to easily digestible, meaningful bits. Classification systems provide a fundamental reference frame in all sciences, “containing the concepts upon which theory is based and influencing what can and can not be seen” (Bogenschutz & Nurnberg 2000). The classification of diseases, or nosology, has always been an integral part of medicine.
Nosology is the process by which medically relevant phenomena are sorted into categories in accordance with certain established criteria, for one or more purposes. Diagnoses are the primary categories of interest, referring to various concepts, ranging from a consistent cluster of signs and symptoms (syndrome) to the effects of a specified aetiological agent.
Classification is necessary to provide a conceptual framework within which to place what is observed, to enable efficient communication about the conditions of illness, to predict outcomes, to guide general treatment decisions, and to keep records. These purposes can be summarised as communication, control, and comprehension. Psychometrically, they are substantiated by the procedures for establishing validity.
Communication involves the use of names of categories as standard shorthand ways of summarising a great deal of information. For communication to be effective, there must be a high level of agreement among the users of a classification system. Control of disorders refers to the application of knowledge about their course with or without intervention, prevention or treatment, and is best supported by high predictability. Comprehension concerns more fundamental aspects of disorders, such as the process(es) involved in their origin, development and maintenance, and requires a high level of convergent

and discriminant validity (Zimmerman & Spitzer 2005). Medical classifications are relatively open systems of hypothetical constructs subject to falsification and further scientific inquiry. Meaningful research questions can be asked about them, which in turn influences the systems themselves.

Despite its essential similarity to classification in the rest of medicine, psychiatric classification is often held as less 'medical' or scientific than it is in other branches of medicine, because mental disorders are generally of a typological nature, with less emphasis on clear boundaries than on the core of each disorder. Frantic attempts to correct this deprecatory judgement resulted in a paradigmatic shift from interpretative to empirically-based approaches, embodied, for example, in the Diagnostic Statistical Manual of Mental Disorders III (DSM-III) and its successors.

Attempts, as in this thesis, to adequately classify mental disorders are at least as old as recorded history. They reflect the ideas of a particular epoch on the criteria and purposes of the classification of abnormal behaviour characterised by deviance, distress, dysfunction, and/or danger. Menninger compiled a compendium of these kinds of classification from ancient times to the modern era (Menninger 1963). From the many topics relevant to these classifications, we will elaborate on the meta-theoretical positions of nominalism versus those of realism, and of empiricism versus those of rationalism, on the theory of classification, and on the operational aspect of the increasing complexity of classification systems.

3.2 Meta-theoretical aspects

Although of major importance, the ontological status of diagnostic categories is hardly ever discussed in psychiatry. The choice is between realism and nominalism. Realism assumes that an entity exists independently of its being named. Nominalism asserts that an entity has no reality independent of its name; diagnostic categories are cultural constructs, not independently existing entities. Since Plato, there has been an ongoing controversy between these two points of view. This centuries-old discussion is still recognisable, for example, in current policy-making papers on the financing of mental health care systems. Some of the authors of these papers consider psychiatric diagnoses as professional labels for problems of living (Kuypers et al. 2002), and for others they serve as real anchor points for reimbursement (managed care) (Maylath et al. 2006; Rushton et al. 2002). Whatever the results of this polemic, psychiatrists, in their roles as clinicians, social agents and scientists, must realise that patients fitting a certain type of diagnosis may be very different from one another in other relevant respects, that the social consequences of belonging to a diagnostic category are hardly evidence-based, that the general public tends to reify diagnostic categories, and finally, that diagnostic categories are in fact man-made typologies.

In the approach to acquiring medical knowledge, there is an epistemological debate about which of two philosophical schools should predominate: empiricism, which stresses experience, or rationalism, which assigns primacy to reason. The debate can be traced back to classical Greece. In that epoch, both Hippocrates and Plato created a system for classifying disorders. Hippocrates developed a system based primarily on empirical observation, whereas Plato's system was rooted in rational idealism. On the island of Kos, the Hippocratic School taught an individualised kind of medicine in which the patient was the primary focus of attention. On Knidos, from which Kos could be seen and which was located on the mainland of Asia Minor, the Platonic endeavours were first and foremost aimed at the elaboration of a nosology (Lindeboom 1971).

After the Middle Ages, in which classification was predominantly a religious matter (punishment for sin or a test of faith), the Renaissance, and especially the Enlightenment, brought a revival of the scientific leitmotiv. Refinement of technical equipment (e.g., the microscope) and the growing independence of medicine from church and state gave rise to the practice of systematic observations and planned experiments, leading to the descriptive era. Thomas Sydenham, and later Carolus Linneaus and François Boisser de Sauvages, attempted to apply the taxonomic methods of biology to medical and psychiatric illnesses, with categories based on observed signs and symptoms. At the end of the 19th century, the apex of empiricist nosology was reached by Kraepelin, who conceived a classification system for psychoses based on extensive

observations in the newly established asylums. On the basis of his experiences with neurotic patients, Freud, on the other hand, developed a (psychodynamic) theory that guided his observations and interventions and that dominated the field of psychiatry for several decades. The influential psychiatric methodologist and philosopher Jaspers, a contemporary of Freud, summarised this approach aptly, as follows:

> *"Die Wirklichkeit wird durch die Brille der Theorie gesehen. Es ist daher ständig unsere Aufgabe, von Theoretische Vorurteilen, die jederzeits in uns wirksam sind, absehen zu lernen uns zu üben, rein die Befunde aufzufassen."* (Jaspers 1973).
> (Reality is always viewed through the prism of theory. It is therefore our unceasing responsibility to find a way of resisting the influence of theoretical preconceptions, which are always active within us, and of applying ourselves to the understanding of observations in their untainted form.)

Even though it eventually became obvious that both empiricism and rationalism were needed for adequate classification, Jaspers' call for unbiased observations was interpreted in terms of a kind of superficial empiricism in the Anglo-Saxon countries. In other words, his phenomenological, or in-depth, version of empiricism according to which only beliefs about one's own sensory experience are directly corroborated was generalised to the idea that beliefs about what we perceive in the physical environment should be directly confirmed by experience. As empiricism re-emerged in the 1950s under the influence of biological and epidemiological tendencies, it therefore took the form of superficial or neo-Kraepelinian empiricism, which was incorporated into what is now known as the DSM tradition. From a rationalistic point of view, however, the DSM-III and its successors are handicapped by the lack of an underlying theory, notably about aetiology.

3.3 Theoretical aspects

3.3.1 Categories versus dimensions

Classifications can be made in terms of categories or dimensions. In medicine (and consequently in psychiatry) the categorical model is traditionally preferred because practically, it implies that a certain disorder is either present or not present, and hence that a treatment should or should not be considered. If a diagnosis has been made for a patient, this means that his or her health differs qualitatively from a conception of a health norm. In a pure categorical system, all diagnostic criteria are necessary and sufficient to make a diagnosis. Patients with a given diagnosis are considered to be a homogeneous group in this respect.

The categorical model, however, is increasingly being criticised. According to Craddock & Owen (2005) for example, the distinction between bipolar affective disorder and schizophrenia made by Kraepelin and accepted to the point where it is enshrined in current classifications, is now reaching the end of its useful life. On the basis of a review of the evidence from both epidemiological and molecular genetics research, they concluded that psychiatry would be better served if a spectrum or a dimension of functional psychoses were conceptualised. In such a dimensional model of classification there are no categories; individuals are described in quantitative terms along continuous factors that have a (more or less normal) distribution throughout the population. Because the dimensions are continuous, various intermediate measures between the two poles of the dimensions can be expressed quantitatively. In a multidimensional classification, individuals can be characterised by their position along a limited number of dimensions. Examples are the two-dimensional system for depression and anxiety proposed by Goldberg and Huxley (1992) and the three-dimensional model for generalised anxiety, phobic anxiety and depression put forward by Ormel et al. (1995).

Be this as it may, pathology in a dimensional system represents a statistical deviation from a quantitatively defined norm. Dimensional approaches are found in the DSM-IV with reference to axes II (personality disorders) and V (social functioning). Both the categorical and the dimensional model have advantages and disadvantages: the categorical model is informative as far as treatment and prognosis are concerned, whereas the dimensional model is better suited for shedding light on the overlap of syndromes and facilitating aetiological research.

3.3.2 Classification versus typology

The fundamental elements of any classification are its commitments to theory, basic units and the criteria for ordering these basic units into a classification. In biological classification, for example, evolution supplies the theoretical orientation. Here, the goal is to make the basic units of classification identical to the basic units of biological evolution. Medical classification or nosology aims at predicting the maximum possible number of relevant facts about diseases. Psychiatric classifications ideally consist of mutually exclusive and jointly exhaustive hypothetical disorders (diagnostic categories) and also rules for making the diagnosis for each category. In a traditional system of classification, categories are defined by a number of individually necessary and jointly sufficient (or monothetic and conjunctive) criteria. So that they can be applied unambiguously, these criteria should be defined as operationally as possible, i.e. as measurable variables. Once the rules defining the category are applied to a case – a process called diagnostic classification, or in short, diagnosis – it can be determined whether or not this case belongs to a given category. Thus the boundaries of the categories are rigorously defined and individuals within such a category are homogenous with respect to the defining characteristics of the category. To meet the demand for mutual exclusiveness of the categories, larger categories of a superordinate level of abstraction are indicated, analogous to the genera, families, etc. in the classification of biological species, and a hierarchy is then imposed on the system. This kind of hierarchical categorical system works to the extent that the relevant classes have such a structure. Unfortunately, like many common object categories, most psychiatric illnesses do not have clear defining features. Consequently, modern psychiatric classifications have been based on typology. One or more, but not all (polythetic and disjunctive) features have to be present for membership of a category. Typology has been called ‘the opposite of true classification’ because it puts similar things together, rather than separating different things from one another (Bogenschutz & Nurnberg 2000). The typological view permits the existence of borderline and heterogeneous cases within a given category. Typal categories have clearly defined centres, but their boundaries are fuzzy. As far as hierarchy is concerned, it is not necessary for members of a subset to have all of the characteristics of the more inclusive set. The recession of the hierarchical approach in the transition from the DSM-II to the DSM-III and its successors has inevitably led to an increase in the prevalence rate of comorbid psychiatric disorders (Kessler et al. 1997).

Because the prototypal model, to a greater extent than the categorical model, is widely regarded as resembling the way in which clinicians conceive of and actually use diagnostic categories in psychiatry, the DSM model in force is basically designed as a prototypal model. This fact has also led to a specification of former, more global categories, resulting in more explicit

separate diagnostic categories such as sleep disorders, somatoform disorders, and dissociative states, so that borders between categories are possibly emphasised where they should not exist.

3.3.3 Classification – its recent history

The demands of the authorities for statistical (census) data created the need for a classification of mental disorders in the 19th century. In the course of the 1800s, the number of categories was increased from one ("idiocy/insanity") to seven in 1880 (DSM IV 1994). In 1883 Kraepelin published his *Kompendium der Psychiatrie*, in which he first presented his nosology or classification of disorders, dividing mental illnesses into exogenous, treatable disorders, and endogenous, incurable disorders (Kraepelin 2007). He continued to refine his classification, issuing nine revisions of his psychiatry textbook, which grew into several volumes. In the sixth edition (1899), he first made the distinction between manic-depressive psychosis and *dementia praecox*, now called schizophrenia. His classification of mental disorders served as the foundation for the versions of the DSM and the International Classification of Diseases (ICD), used by psychiatrists today.

Between 1920 and 1940, health economics, the increasing role of government in healthcare, and the military greatly facilitated the development of an unambiguous taxonomy of diseases, including mental diseases. The first international classification of mental diseases appeared in the sixth edition of the ICD (1948), published by the WHO. Because the mental diseases section of the ICD-6 was not found to be satisfactory in the United States, in 1951 its Public Health Service commissioned a committee to draw up an alternative to the mental diseases section contained in ICD-6, which resulted in the DSM-I in 1952. Despite its significant impact and influence on American psychiatric literature, the DSM-I was not universally accepted as an official nomenclature. Although the WHO promoted an international effort to develop a system of classification for mental diseases which would improve the ICD-6 section and meet the requirements of all member nations, the two classification systems have co-existed and been developed side by side ever since. In ICD-8, individual syndromes and diseases are collected in homogeneous categories and organised in a mutually exclusive, hierarchical fashion. At the same time (1968), the new diagnostic manual, i.e. the DSM-II, was published, which was compatible with the list of mental diseases in ICD-8 but adapted for use in the United States. The DSM II was adopted by the American Psychiatric Association and officially accepted nationwide. The ICD-9 and the DSM-III were introduced in the early 1980s. The latter, which was published in 1980, brought in significant methodological innovations, such as a diagnostic algorithm based on clear defined criteria with rules for inclusion and exclusion, the system of the

multi-axial classification and the declared neutral attitude regarding the explicative theories, making it exclusively descriptive.
Subsequent sharpening of the diagnostic criteria led to the publication of the DSM-III-R (1987) and the DSM-IV (1994), parallel to the WHO publication of the ICD-10 in 1992. Their respective nosographies have become more homogeneous and, consequently, more comparable.
At present the ICD-10 is widely used in Europe. In the United States, however, the changeover to the ICD-10 was complicated by the fact that the ICD-9-CM (Clinical Modification) was part of the hospital billing system. The U.S. National Center for Health Statistics has set a deadline, as recommended by the National Committee on Vital and Health Statistics, for the adoption of the current revision of ICD-10-CM, which was pre-released in autumn 2003.
Parallel to this evolution, modern psychopharmacology and genetic research continue the search for the core of psychopathological processes with ever-more powerful investigational instruments.

3.3.4 Classification – advantages and disadvantages

Given the potential clues for treatment offered by a diagnostic classification, we must not forget that assigning patients to diagnostic categories has limitations and potential ill effects.
Attitudes to psychiatric classification have undergone a revolution in the past few decades. From the 1950s to the 1970s, psychiatric diagnoses were held in low esteem and some of the rationales of this judgement can still be heard today. Psychiatric diagnoses were, and sometimes still are, considered a very inadequate means of conveying what the clinician believes to be the essence of a patient's misery. Also, many psychiatric diagnoses have pejorative connotations, which might prejudice the attitude of other people to the patient and so the attitude of the patient to him- or herself. Furthermore, attaching a name to a condition may give rise to a spurious sense of understanding; some clinicians reify a diagnostic concept and treat the disease instead of trying to relieve their patients' burden. The greatest danger posed by classifications is the potential reification of hypothetical approaches, arbitrary categorisation and the dulling of reflection, all of which have created a need for regular revisions underpinned by field trials (Lemperiere 1995). Furthermore, before the 1980s, the reliability of psychiatric classification was known to be low (Beck 1962; Ward et al. 1962a), and, as confirmed by the US-UK study, key diagnostic terms like schizophrenia had different meanings in different parts of the world (Andreasen 1989).
In the 1980s the attitude towards psychiatric classification changed radically. This shift might be partially explained by the rise of the biologically (at the expense of the sociologically) oriented approach to deviant behaviour, and

partially by the attempts of psychiatry to improve the scientific foundation of its classification systems. On balance, this resulted in the development and worldwide acceptance of the DSM-III and its successors.
This remarkable change of attitude towards classification reflects the transition from emphasis on the shortcomings to emphasis on the strengths of the classificatory system. These strengths and shortcomings are intrinsically related to the essential purpose of classification: to distinguish between characteristics of patients. Every patient possesses characteristics of three kinds (Kendell 1975b; Kendell 1993):

- those, he/she shares with all other patients;
- those, he/she shares with some other patients, but not all;
- those which are unique to him/her.

In so far as the first of these three categories is dominant, classification is pointless, since it results in just a single category. To the extent that the third category is dominant, classification is impossible, as every category contains just one individual. As soon as one begins to recognise features that apply to some patients but not to all, and to distinguish those that are important from those that are not, one is classifying them. If we want to use different treatment strategies with maximum efficacy, we have no alternative but to distinguish between one type of patient and another. A distinction between different kinds of mental disorders, consequently, is inevitable in any situation in which groups of patients need to be considered. As classifications of mental disorders are still largely based on differences in symptomatology rather than on differences in aetiology, we must try to improve the classification systems we possess. A future classification should be more transparent, stable, and valid (Kendell 1993).
Another important but often disregarded question we would like to put is whether a classification system can be used for the general population or is it meant only for clinical (sub)groups. This question has to do with the basic problem of defining ‘caseness’. After all, traditional psychiatric classifications are based on symptoms of patients encountered in psychiatric hospitals, and one may ask whether these symptoms are representative of individuals in the general population.

3.3.5 The psychopathological spectrum

As mentioned previously, the central pillar of epidemiology is the condition, which has to be defined unambiguously in a case definition. This central question of ‘what is a case’ (diagnosis) is as old as mankind. As a major need even in ancient medical practice, case definition has undergone transitions from simplicity, resulting in a dichotomy (healthy and ill or believer and doomed) to a system in which biological, psychological, sociological, religious and political

screening filters, under the influence of technical refinement and treatment potential, play a role, that has resulted in a set of hundreds of more or less distinguishable disease entities. This situation has of course given rise to the phenomenon of more than one diagnosis for a given person at a given time (comorbidity) (First 2002).

The art of healing (and consequently of diagnosis) is one of the oldest intellectual attainments of human beings and it arose from human limitations, need, self-protection, and the urge to help one's fellow human being (Magner 1992). The ancient Greeks established intellectual traditions that provided the foundations of Western philosophy, science and empirical medicine. In contrast, for example, to the ancient Indian, Chinese, and Egyptian civilisations, the Greeks managed to separate medicine from religion and placed disease on a far more rational basis. For more than one thousand years after this period, no important new insights into medicine were developed, and the psychopathological spectrum remained restricted to the major entities of psychosis and depression. In the Middle Ages, these were perceived as a punishment for sin or a test of faith and diagnosis was therefore predominantly a religious matter. The Renaissance, and especially the era of enlightenment, brought about veritable revolutions in the understanding of the origin of diseases, but this did not result in an adjustment of the psychopathological spectrum. It was not until the mid-19th century, with the development of pharmacology, cell pathology and bacteriology (Pasteur), that organic syndromes were added to the psychopathological spectrum, with *dementia paralytica* as the most elaborate example. Under the influence of psychoanalysis, this was followed at the beginning of the 20th century by a broadening of the scope to include neurotic and reactive states.

In the last decades of the 20th century, the spectrum was updated with case definitions of dissociative states, various sleep disorders and somatoform disorders.

This expansion of the range of psychopathology can be logically explained in terms of the shift of emphasis in the diagnostic process. Formerly, the use of specific diagnoses had been exclusively dependent on the clinical impressions of authorities in the clinical field (face validity). This practice changed in the last decades of the 20th century, since in response to the demand for transparency and public accountability, there was a requirement for diagnoses to be defined as operationally as possible, i.e. they were to be measurable.

Another explanation for the extension of psychopathology is that the clinical view is not unambiguous. In the 1950s, making a specific diagnosis was the domain of a clinical specialist, with his or her high-powered credentials, but narrow view. In addition, these specialists worked almost exclusively with in-patients. Today, general practitioners are the gatekeepers of specialised (mental) health care, and the vast majority of psychiatrists work with out-

patients. Both of these factors have resulted in a different frame of reference from that of the clinical psychiatrist of the 1950s. The GP now sees a different category of patients, who present sub-clinical manifestations or precursors of the “real” diseases, and he or she applies correspondingly different professional standards in dealing with them. The above-mentioned requirement for transparency and public accountability also resulted in an emphasis on prevention and primary care when possible, whereby the importance of extending the scope of what is defined as ‘not normal’ was stressed.

3.4 Operational aspects

3.4.1 Operational methods of diagnosis

In practice, disease categories are inseparable from the means of diagnosing them, and the diagnostic algorithm should reflect the 'true' structure of the disorder as much as possible. The optimal diagnostic algorithm would be achieved through the use of an infallible criterion that is pathognomonic for the disorder in question. Down's syndrome, for instance, can be diagnosed unequivocally by a karyotypal demonstration of the presence of all or part of an extra 21st chromosome (trisomy 21). Unfortunately, in psychiatry, there are hardly any true gold standards. Consequently, categorical diagnoses concerning mental disorders are made with the aid of a number of diagnostic criteria, each of which is thought to be correlated with the disorder, but none of which is necessary or sufficient to conclude that the individual in question has the disorder. The diagnostic algorithm in psychiatry proceeds from signs and symptoms through syndromes to disorders and diseases. Signs and symptoms are the basic units and syndromes are sets of symptoms that co-occur at a greater than chance frequency. A disorder is the manifestation of the conjunction of a syndrome with a clinical course, although the underlying causes remain obscure. In psychopathology, the term disease is reserved for situations in which signs and symptoms, pathology, underlying causes, and their connections are known. Progress in understanding pathology is like an archaeological process in that the 'digging' proceeds from the superficial signs and symptoms to the underlying disease. A precise understanding of syndrome and pathology course furthers the discovery of mechanisms and causes. The separation of Down's syndrome from the more general category of mental deficiency, for instance, facilitated Lejeune's discovery of the underlying aetiology – trisomy 21 (Lejeune et al. 1959). This kind of constructive distinction is difficult to make in psychiatry because of the prototypal character of the categorical system, in which similarities are looked for instead of differences.

Until the 1950s, the most widely used method in the process of diagnosis was the free-ranging interview, like the classical psychiatric examination. This method is prone to a number of typical errors (De Bruyn et al. 2003; Kreitman et al. 1961; Kreitman 1961), of which two examples follow: first, although people (whether of not professionals) are able to adequately estimate the difference between certain (psycho)physical stimuli (McDowell & Newell 1996), in general, they perform poorly when it comes to assessing, considering, or revising their estimations; second, in the daily judgements that people make, often they do not use reason in strict accordance with logic, or else they leave relevant information out, or make use of irrelevant information. Even when someone has

all the relevant information that is available, the process of modification of estimations in the light of the new material does not proceed optimally.
It is for these reasons that the information gathered with the free-ranging interview, though clinically relevant, is too much influenced by subjective factors and therefore uncommunicable. This fact limits scientific usefulness, which is unacceptable to the psychiatrist-epidemiologist, about whom Cooper said that he or she should choose (diagnostic) variables that are not only clinically, but also scientifically, relevant (Cooper 1979). Diagnostic instruments in psychiatry ought to be diagnostic algorithms that have been constructed in such a way that they are transparently operational. The recognition of the aforementioned shortcomings, together with the improved classification systems, encouraged the development of a diagnostic instrument with better psychometric properties, without detriment to the clinical perspective.

3.4.2 Diagnostic instruments

With a view to making the process of classifying mental disorders operational, various instruments have been designed in the past few decades. The two most important general diagnostic instruments today are the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (Janca et al. 1994; Wing et al. 1990; Wing 1996), which includes the tenth version of the Present State Examination (PSE-10), and the Composite International Diagnostic Interview (CIDI) (Robins et al. 1988), the successor to the Diagnostic Interview Schedule (DIS). In contrast to their predecessors, both the PSE-10 and the CIDI can be used for a diagnostic classification with reference to the DSM-IV and ICD-10; a computer programme generates diagnoses from the item scores, in a consistent way.
The two essential differences between the SCAN (and the PSE), on the one hand, and the CIDI (and the DIS), on the other, are the semi-structured character, including clinical judgments, and the bottom-up approach of the SCAN/PSE versus the fully standardised character without clinical judgment and the top-down approach of the CIDI/DIS (Ustun & Tien 1995). In the SCAN/PSE tradition, the interviewer keeps asking questions about a specific symptom until he or she is satisfied with the answer as to its presence or absence and its severity, without consideration of a possible diagnosis. Only a trained clinician can perform this kind of ‘cross-examination’, and he or she does so on the basis of an extensive manual, in which the psychopathological concepts are clarified and criteria are defined for the presence of symptoms. These criteria rest on three underlying notions about psychiatric symptoms: the patient is suffering from the symptom; the severity of the symptom is not proportionate to the circumstances in which it occurs, and the patient is *not in control* of his or her symptom, i.e. she is not able to put it out of her mind. Fundamental for the SCAN/PSE tradition, consequently, is the symptom-based

approach, which corresponds to the definition formulated by Leighton: Behaviour of Psychiatric Interest', (which) can be seen as a set of basic units out of which the recognised syndromes are assembled, much as the notes of the music scale are combined to make melodies (Leighton 1979). In contrast to the SCAN/PSE, clinical experience is not necessary to administer the CIDI/DIS. Clinical judgment in the CIDI/DIS is replaced by questions concerning help-seeking behaviour of the patient, medication and hindrance. Furthermore, the CIDI/DIS checks for the presence or absence of syndromes as a whole (as classified by the DSM or ICD system), while the SCAN/PSE checks for each individual symptom. These differences favour the CIDI interview in that it is less expensive than the SCAN interview, and consequently the CIDI is the epidemiological instrument of choice for large general population surveys.

As low-cost alternatives for interviews, there are numerous self-administered questionnaires, for example the well-known General Health Questionnaire (Furer et al. 1995b; Goldberg & Williams 1988; Koeter & Ormel 1991b). Although questionnaires result in an assessment of severity, albeit rough, and in a limited diagnostic domain with certain questionnaires, they can not be used to establish signs and symptoms and syndromes, let alone a diagnosis. Some questionnaires approximate caseness, for example when a cut-off point is used.

Each approach has its pros and cons. Symptom-based instruments, such as the SCAN/PSE, have two advantages: their nature is such that they can survive changes in diagnostic classifications, and they are quite comparable to the clinical process(es) involved in diagnosis-making, in that they compensate for the patient's lack of illness awareness. A disadvantage is that the interview is costly, since it can be performed only by well-trained professionals and clinicians and it is time-consuming (on average it takes about 90 minutes to administer). Diagnostic instruments for epidemiologic surveys, like the CIDI/DIS, have the advantage that they are suitable for use in large-scale surveys since they are cost-effective. The cost saving arises from the fact that lay interviewers are used and these instruments take on average only about 60 minutes to administer. There are two disadvantages to these fully standardised approaches: first, not all positive ratings necessarily denote the presence of a clinical symptom, but rather the subjective experience of a malaise, and second, there is a risk of false negatives with respect to symptom identification because of their top-down structure, or the subject's lack of illness awareness. In large-scale surveys, when underlying relations are investigated, false positives and negatives can be seen as noise, on the assumption that there is no biasing systematic effect. Finally, a major disadvantage of general questionnaires is that a diagnosis can not be made on the basis of the results obtained from them. Their strength lies in the fact that they can be used as a convenient and cheap means of screening people at risk from mental disorders in general population surveys.

3.5 Psychometric evaluation of diagnostic procedures

The main question before us now is, do current standardised psychiatric instruments successfully reflect an explicit and accepted definition of mental ill health. The issue here is validity: the assessment of whether a measurement really measures what it is supposed to. This matter is particularly important, as in real life, people have a personal stake in the estimation of their mental health. Someone who is learning to shoot must first learn to hit the centre of the target, and then learn to do this consistently. This metaphor is analogous to the situation regarding the validity and reliability of a diagnostic procedure or diagnostic measurement (McDowell & Newell 1996). The consistency (or reliability) of a diagnostic procedure can be compared to the closeness of successive shots to one other in the target area. Validity can be likened to the target that is being aimed at – the closeness of the shots to bull's eye. Ideally the shots should cluster in the centre of the target (reliable and valid), but for a marksman who is consistently off target they would cluster at a distance from the centre (reliable but not valid).

A third psychometric quality of instruments is their responsiveness or sensitivity to changes in the state of the subjects of interest, described as "an instrument's ability to detect clinically important changes in patient status" (Deyo & Patrick 1989). The SCAN/PSE is based on an assessment of the clinical relevance of symptoms. Since the SCAN/PSE is relatively independent of the classification system in use, it can be expected to incorporate this capability. Since measurement of change (from 1983 to 1997) was one of the goals of this study, we will elaborate on this item.

3.5.1 Assessment of validity

Validity of measurement is an even greater concern than reliability of measurement. After all, what is the good of being consistently wrong? (Gould 2002; McDowell & Newell 1996) Health and illness are complex phenomena that are not directly accessible to measurement. All health measurement instruments can be looked upon as indicators of hypothetical health constructs embedded in a theoretical network of other health-related constructs (König-Zahn et al. 1993). In other words, as Carmines and Zeller stated (1979), (health) measurement is a process of linking unobservable theoretical concepts to empirical indicators. According to a broad definition (McDowell & Newell 1996), validity describes the range of interpretations that can be appropriately put on a measurement score: what can we conclude about a person with a particular score on a measurement instrument? This general definition has advantages and disadvantages as compared with the commonly used more specific definition of validity as the extent to which an instrument measures what it is intended to measure. The broad definition may stimulate a search for

interpretations of an indicator outside the intended scope of the instrument, which may lead to questions about alternative possibilities, and hence to the discovery of links between constructs that were previously thought to be independent. It may also yield valuable insights into the scope of instruments. Dementia screening tests, for instance, are valid in that they succeed in their purpose of identifying cognitive impairments, but they also appear to detect people with little education, which suggests that the tests are a less specific indicator of cognitive functioning. The validity of an instrument should always be measured against its purpose (Carmines & Zeller 1979). If we want to know the prevalence of treatable diseases, we should use instruments that were devised from a clinical perspective; however, a questionnaire that is internally valid might generate results that at first glance are similar to those obtained from such instruments. A disadvantage of broadening the definition of validity is that that could give rise to carelessness in defining the precise purpose of a measurement. Because of this, much time has been wasted, for example, on speculation about what certain psychological well-being scales are actually supposed to measure. This pitfall can best be avoided through closely linking the validation process to a conceptual formulation of the aims of measurement and also linking such a formulation to other, related concepts so that the possibility of alternative interpretations of scores can be taken into account.

The three most frequently used methods of validity testing, i.e. construct, criterion, and content validity, will be discussed briefly. Face validity is not considered by most experts as a true type of validity. In the case of a measuring instrument, most authorities in the field view it instead as giving a first impression of the comprehensibility, the clarity and the appropriateness of the questions of the instrument.

3.5.1.1 Construct validity

The crucial feature of scientific research is the measurement of abstract concepts (constructs) and their relationships to other abstract concepts. From this perspective, validity can be described as the degree to which the score obtained from an instrument, which is an empirical indicator for a theoretical concept, actually represents the concept of interest. In fact, of all three types of validity, construct validity is all-embracing and the most complex, and is therefore a never-ending process. With construct validity, two different types of testing are conducted – convergent and discriminant validity testing. The first tests whether the outcomes of the instrument under study are highly correlated with measured values of indicators of equal usefulness for the same health construct. In contrast, the second tests whether there is a low correlation with measured values of indicators used for different health constructs.

3.5.1.2 Criterion validity

Criterion validity asks whether the measure compares well with external standard measures (so-called gold standards). There are two subtypes of criterion measures, the predictive and the concurrent measure. The first asks whether the test accurately predicts behaviour in the true situation, e.g. does the ability to learn new or uncommon words predict the ability to learn foreign languages at school? The validity coefficient is a measure of the relationship between scores on the prediction test and on actual performance in the criterion situation (Gould 2002). The second type, concurrent validity, measures the correlation between the test under study and another assessment of the situation carried out at the same time, e.g. the correlation of a new test of depression and a well-known valid assessment tool, i.e. a gold standard.

3.5.1.3 Content validity

With content validity the key question is, how representative are the test items of the content of the health construct. There are two standards by which content validity is assessed: the representativeness of the collection of items chosen and the type of test construction used to measure the concept. Content validity is considered to be less objective than criterion validity. There is no specific statistical means of testing content validity, and so reason or else a consensus among experts must be relied on as regards the representativeness of the content.

3.5.2 Assessment of reliability

Reliability is defined as the consistency or stability of the measurement process across time, respondents or observers (Gould 2002), and should not be confused with a term like trustworthiness. Although consistency of measurement describes the phenomenon in a more unambiguous way, reliability is the most widely accepted term, and consequently is used here.

There are many sources of measurement error, and, put simply, each measure is the summation of the true score, the constant (or systematic) error, and the variable (or random) error. Systematic errors (or biases) are more a validity then a reliability issue, and are therefore generally considered in the context of validity testing.

Random errors are inaccuracies that may, in equal measure, lead to an over- or underestimation. Random error is formally defined as the proportion of observed variation in scores that is equally distributed over the low and the high side of the true score. It may be due to inattention or fatigue on the part of the rater or mechanical inaccuracy of the test instrument. McDowell and Newell (1996) cite three different types of reliability testing. The first type tests whether

different raters assessing the same respondent obtain the same result (inter-rater agreement). Such agreement often leads to an overestimation of reliability, for example when one person performs the interview and also does the scoring, while a second person only rates the answers to the questions put in the interview. In this case, the differences in interview techniques are eliminated. This disadvantage can be overcome by a second type of testing, namely having the respondent re-interviewed after a brief interval by each of the raters separately. In this test-retest procedure, reliability is underestimated as there is a difference in time between the two interviews. Moreover, the respondent might react differently during the second interview, resulting in a lower reliability score, if only because the same test is being performed twice. The same applies to the third type of test, the intra-rater reliability test, which assesses whether the same result is obtained when the same rater makes a second assessment of the same respondent (McDowell & Newell 1996).
Gould (2002) adds more sophisticated techniques to the aforementioned tests, such as the alternate-forms technique (calculation of the correlation between two successive measures on two different, but equivalent, versions of the same test) and the split-half technique (calculation of the correlation between two simultaneously obtained measures from equivalent halves of a single test). Combinations of the aforementioned techniques are often used in testing the reliability of new interviews and questionnaires.
Kendell addresses three methodological principles for improving reliability (Kendell 1993):

1. structured and standardised instruments should be used;
2. all symptoms and signs that are determined should be well defined in such a way that the rater understands the definitions;
3. any algorithms that are used should be transparent and unambiguous.

It goes without saying that the way in which an instrument is presented or used is crucially important to its reliability. When interviewers perform a semi-structured interview, as in our study for example, they should be experienced not only in interview techniques, but also in the psychopathology to be determined.
Another way, in fact the most commonly used way, of improving reliability is to create a high internal consistency among the set of items to be measured. Such consistency is a powerful property that manifests itself in increased repeatability and consequently increased reliability. But as this factor often reduces the scope of an instrument, it mostly improves reliability at the cost of a decreased clinical validity and also of a lower content validity. Hence there is an equilibrium between validity and reliability; to be valid, a measurement must be highly

reliable, but too high a reliability score as a result of an extremely high internal consistency, on the contrary, might compromise the validity.

3.5.3 Sensitivity to change

Sensitivity to change (or responsiveness) refers to the change scores obtained with a health status instrument, as if it were a screening test for the detection of true change, e.g. in the measurement of responsiveness to a certain treatment(Deyo et al. 1991; Deyo & Inui 1984; Deyo & Patrick 1989; Fitzpatrick et al. 1998). Guyatt (1989) defined responsiveness as the ability of an instrument to detect clinically important change. In a strict sense, this procedure is meant to measure the sensitivity to change over time in a specified group or cohort. Generally speaking, scale data are preferred over ordinal data, and both are preferred over frequency data because of their greater precision of measurement and hence sensitivity (Gould 2002). From this point of view, caseness as a measurement result from the SCAN/PSE is not a sensitive measure for assessing change, as it is binominal in character. On the other hand, in many psychiatric epidemiological studies, measurement of caseness has been used in the general population, with frequency data as a result. Together with the large numbers of respondents, this approach generates sufficient power for analysis.

Finally, it is important to realise that most diagnostic instruments (like the SCAN and the CIDI) measure over a 1-month reference period for most disorders, but not for example, for somatisation and disorders related to substance use, which require a much longer reference period. Consequently, in this regard, sensitivity to change for the two last-named disorders is different from what it is for, for instance, depressive disorder, especially when the time interval between the two surveys is relatively short in relation to the reference period. This means that a protracted interval in a repeated cross-sectional survey conduces to a better sensitivity to change than a short interval of just 1 or 2 years.

3.6 Conclusions

In this study, the Nijmegen Health Area 2 project (NHA-2), we chose to use the SCAN-2.1 as an instrument in the PSE tradition for the following reasons: first, we believed it would enhance the comparison with the data obtained in the first project, in 1983; second, it fulfilled the requirement that the psychiatrist-epidemiologist should choose his or her instruments and variables not only for their clinical, but also for their scientific, usefulness (Cooper 1979). Furthermore, since in real life people have a personal stake in the estimation of their health, the SCAN/PSE, with its observer-based approach, provided us with a professionally oriented assessment. referral

The SCAN-2.1 covers almost the total psychopathological spectrum and its psychometric properties are highly acceptable (Andrews et al. 1995; Tomov & Nikolov 1990), as will also be proved in this thesis. Finally, in the Netherlands there is a long-standing SCAN/PSE tradition, and associated with it, authorised training facilities.

References

Andreasen NC. (1989) The American concept of schizophrenia. *Schizophrenia Bulletin,* 15, 519-531.

Andrews G, Peters L, Guzman AM and Bird K. (1995) A comparison of two structured diagnostic interviews: CIDI and SCAN. *Australian and New Zealand Journal of Psychiatry,* 29, 124-132.

Beck AT. (1962) Reliability of psychiatric diagnosis: a critique of systematic studies. *American Journal of Psychiatry,* 119, 210-216.

Bogenschutz MP and Nurnberg HG. (2000) Classification of mental disorders. In BJ Sadock en VA Sadock (Eds.), *Kaplan and Sadock's Comprehensive Textbook of Psychiatry*. ed. 7 pp. 824-830. Lippincott Williams & Wilkins, Philadelphia.

Boot JMD and Knapen MHJM. (2001) Population and health [in Dutch]. In *Handbook of Dutch Health Care*. ed. 7 pp. 17-50. Het Spectrum, Utrecht.

Carmines EG and Zeller RA. (1979) *Reliability and validity assessment. Quantitative applications in the social sciences*. Sara Miller McCune, SAGE publications, inc., Beverly Hills / London.

Cooper B. (1979) *Demographic and epidemiological methods in psychiatric research*. Springer Verlag, Berlin.

Craddock N and Owen MJ. (2005) The beginning of the end for the Kraepelinian dichotomy. *British Journal of Psychiatry,* 186, 364-366.

De Bruyn EEJ, Ruijssenaars AJJM, Pameijer NK and van Aarle EJM. (2003) De diagnostische cyclus. In EEJ De Bruyn, AJJM Ruijssenaars, NK Pameijer en EJM van Aarle (Eds.), *De diagnostische cyclus*. pp. 31-49. Acco, Leuven.

Deyo RA, Diehr P and Patrick DL. (1991) Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Controlled Clinical Trials,* 12, 142S-158S.

Deyo RA and Inui TS. (1984) Toward clinical applications of health status measures: sensitivity of scales to clinically important changes. *Health Services Research,* 19, 275-289.

Deyo RA and Patrick DL. (1989) Barriers to the use of health status measures in clinical investigation, patient care, and policy research. *Medical Care,* 27, 254-268.

DSM IV. (1994) *Diagnostic and Statistical Manual of Mental Disorders; Fourth edition*. American Psychiatric Assiciation, Washington, DC.

First MB. (2002) DSM-IV and Psychiatric Epidemiology. In MT Tsuang en M Tohen (Eds.), *Textbook in psychiatric epidemiology, second edition*. pp. 333-342. Wiley-Liss, New York.

Fitzpatrick R, Davey C, Buxton MJ and Jones DR. (1998) Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment,* 2, i-74.

Furer JW, König-Zahn C and Tax B. (1995) General Health Questionnaire (GHQ). [in Dutch]. In JW Furer, C König-Zahn en B Tax (Eds.), *The measurement of health*. pp. 104-118. van Gorcum, Assen.

Goldberg DP and Huxley P. (1992) *Common mental disorders: a biosocial model*. Tavistock/Routledge, London.

Goldberg DP and Williams P. (1988) *A user's guide to the General Health Questionnaire*. NFER-Nelson Publishing Company, Windsor.

Gould JE. (2002) *Concise handbook of experimental methods for the behavioral and biological sciences*. CRC Press, Boca Raton, Florida, USA.

Guyatt GH, Deyo RA, Charlson M, Levine MN and Mitchell A. (1989) Responsiveness and validity in health status measurement: a clarification. *Journal of Clinical Epidemiology,* 42, 403-408.

Janca A, Üstün TB and Sartorius N. (1994) New versions of World Health Organization instruments for the assessment of mental disorders. *Acta Psychiatrica Scandinavica,* 90, 73-83.

Jaspers K. (1973) *Allgemeine Psychopathologie*. (9 ed.) Springer Verlag, Berlin.

Kendell RE. (1975) *The role of diagnosis in psychiatry*. Blackwell Scientific Publications, Oxford, UK.

Kendell RE. (1993) Diagnosis and classification. In RE Kendell en AK Zealley (Eds.), *Companion to psychiatric studies*. ed. 5 pp. 277-294. Churchill Livingstone, Edinburgh, UK.

Kessler RC, Anthony JC, Blazer DG, Bromet E, Eaton WW, Kendler K, Swartz M, Wittchen HU and Zhao S. (1997) The US National Comorbidity Survey: overview and future directions [editorial]. *Epidemiologia e Psichiatria Sociale,* 6, 4-16.

Koeter MW and Ormel J. (1991) *General Health Questionnaire: manual [in Dutch]*. Swets and Zeitlinger, Lisse.

König-Zahn C, Furer JW and Tax B. (1993) *The measurement of health [in Dutch]*. (Vols. 3) v. Gorcum, Assen.

Kraepelin E. (2007) *Kompendium der Psychiatrie. Zum Gebrauch für Studierende und Ärzte*. Vdm Verlag Dr. Müller, Saarbrücken.

Kreitman N. (1961) The reliability of psychiatric diagnosis. *Journal of Mental Science,* 107, 876-886.

Kreitman N, Sainsbury P, Morrissey J, Towers J and Scriverner J. (1961) The reliability of psychiatric assessment: an analysis. *Journal of Mental Science,* 107, 887-908.

Kuypers P, van Londen J, Hutschemaekers G, Duyvendak JW, Schnabel P, Visser ASC and de Winter M. (2002) *Zorg van velen. Eindrapport van de Landelijke Commissie Geestelijke Volksgezondheid [in Dutch]*. Ministerie VWS, Den Haag.

Leighton AH. (1979) Research directions in psychiatric epidemiology. *Psychological Medicine,* 9, 235-247.

Lejeune J, Turpin R and Gautier M. (1959) Chromosomatic diagnosis of mongolism [in French]. *Archives francaises de pediatrie,* 16, 962-963.

Lemperiere T. (1995) The importance of classifications in psychiatry [in French]. *Encephale,* 21 Spec No 5, 3-7.

Lindeboom GA. (1971) *Introduction into the history of medicine [in Dutch]*. De erven F. Bohn, Haarlem.

Magner LN. (1992) *A history of medicine*. Marcel Dekker Inc., New York.

Maylath E, Krokotsch A and Kunze H. (2006) For and against: diagnosis related groups in psychiatry? [in German]. *Psychiatrische Praxis,* 33, 56-58.

McDowell I and Newell C. (1996) *Measuring Health; a guide to rating scales and questionnaires*. (2nd. ed.) Oxford University Press, New York.

Menninger K. (1963) *The vital balance: the life process in mental health and illness*. Viking Press, New York.

Nurcombe B and Gallagher RMI. (1986) Patterns: categorical diagnosis and DSM-III. In B Nurcombe en RMI Gallagher (Eds.), *The Clinical Process in Psychiatry: Diagnosis and Management Planning*. pp. 225-242. Cambridge University Press, Cambridge.

Ormel J, Oldehinkel AJ, Goldberg DP, Hodiamont PP, Wilmink FW and Bridges K. (1995) The structure of common psychiatric symptoms: how many dimensions of neurosis? *Psychological Medicine,* 25, 521-530.

Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA and others. (1988) The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry,* 45, 1069-1077.

Rushton JL, Felt BT and Roberts MW. (2002) Coding of pediatric behavioral and mental disorders. *Pediatrics,* 110, 167-168.

Semple D, Smith R, Burns G, Darjee R and McIntosh A. (2005) *Oxford Handbook of Psychiatry*. Oxford University Press, Oxford.

Tomov T and Nikolov V. (1990) Reliability of SCAN categories and scores. Results of the field trials. In CN Stefanis (Ed.), *Psychiatry: a world perspective*. pp. 107-117. Elsevier Science Publishers, Amsterdam.

Ustun TB and Tien AY. (1995) Recent developments for diagnostic measures in psychiatry. *Epidemiologic Reviews,* 17, 210-220.

Ward CH, Beck AT and Mendelson M. (1962) The psychiatric nomenclature: Reasons for diagnostic disagreement. *Archives of General Psychiatry,* 139, 138-143.

Wing JK. (1996) SCAN and the PSE tradition. *Social Psychiatry and Psychiatric Epidemiology,* 31, 50-54.

Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D and Sartorius N. (1990) SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry,* 47, 589-593.

Zimmerman M and Spitzer RL. (2005) Classification in psychiatry. In BJ Sadock en VA Sadock (Eds.), *Kaplan and Sadock's Comprehensive Textbook of Psychiatry*. ed. 8 pp. 1003-1052. Lippincott Williams & Wilkins, Philadelphia.

4 Research objectives

4.1 Introduction: case finding and case identification

To summarise the theoretical background of chapters 2 and 3, although the application of epidemiological methods in psychiatry is a relatively recent development, it is of scientific importance in exploring the prevalence rate and distribution of psychiatric disorders, improving etiological understanding and enhancing treatment, prevention and planning in psychiatry. Whatever the objective, it is essential that prevalence rates should be not only reliable, but also valid; in other words, prevalence rates must be clinically relevant.

Psychiatric epidemiology turns around the question of what condition occurs in whom, where and when. If this question is to be answered properly, the whole complex of case definition, case identification and case finding has to be taken into account. In the previous chapters, we discussed in detail the various aspects of the principles that subtend these phenomena. The following definitions apply in the rest of this thesis:

- **case definition**: the theoretical concept underlying 'what is a case', based on the presence of a specified set of signs and symptoms. Case definitions are ordered in a classification system, on the one hand by describing the boundaries of each case and thus the distinctness of each case with respect to all the others, and on the other hand, by sorting cases into categories through matching characteristics;
- **case identification**: the operationalisation of a theoretical diagnostic concept or classification system on the basis of a detailed description;
- **case finding**: the application of case identification in a population study.

Epidemiology can contribute to improved communication, control, and comprehension only if case definition is valid and unambiguous, if case identification is based on clinical relevancy and if the process of case finding is reliable and valid. Notwithstanding, reliability and validity in psychiatric case finding have been and still are called into question. Prevalence rates are often not as important or clinically relevant as they seem to be at first glance. Although questionnaires and fully standardised interview instruments are frequently used for case detection in general population surveys, questions can be raised about their clinical validity. On the other hand, clinical interview instruments are time-consuming and expensive when used on a large scale and are often said to lack reliability, although the results of several studies challenge this claim.

4.2 Descriptive epidemiology, i.e. case finding: the prevalence and distribution of identified cases

Many of the subjects raised in the preceding theoretical chapters have played a role in the decisions taken with reference to the design of both the Nijmegen Health Area-1 (NHA-1) and the NHA-2 study. Both of these are descriptive epidemiological case finding studies in the general population of the Dutch Health Area of Nijmegen and they are harnessed to one another in a repeated cross-sectional design. In deciding upon the interview instrument to be used in these studies, sitting on the shoulders of Jaspers, we followed the empirically based, phenomenological clinical tradition. This states that observations should be unbiased, which calls for a theory-free, bottom-up approach to the assessment of symptoms, rather than the assessment of symptoms within the framework of a specific psychiatric syndrome or classification system. Because the ninth version of the Present State Examination interview (PSE-9) had been used for the NHA-1, for the NHA-2 it was decided to use the Schedules for Clinical Assessment in Neuropsychiatry (SCAN-2.1), as the latter instrument is the successor to the PSE-9 and contains the PSE-10 interview. The use of the SCAN-2.1 gave us the opportunity to convert the scores of the latter so that it fitted into the format of the PSE-9 interview algorithm. This conversion made possible a meaningful comparison between the years 1983 and 1997 with reference to the PSE-9 data, as an example of monitoring (see chapter 5), in addition to the comparison with reference to the more general GHQ data (see chapter 11.1).

Thanks to its bottom-up approach to the evaluation of psychiatric signs and symptoms, the SCAN-2.1 made it possible to assess the supplemental epidemiological value of three additional diagnostic categories relatively newly determined in the general population: dissociative disorders, sleep disorders and somatoform disorders (see chapter 6).

4.3 Diagnostics: case identification

As the WHO had recently designed the SCAN-2.1, we had to establish the psychometric properties of the SCAN interview instrument before using it in our field study (see chapter 7).

One of the advantages of the SCAN-2.1 is that in addition to the registration of the absence of a sign or symptom ("0"), the scoring algorithm allows a clinical interpretation of each sign or symptom that is present; specifically, a sub-clinical presence can be scored as "1" and a clinical presence as "2" or "3". The qualification sub-clinical ("1") in particular made it possible to study the properties of the SCAN-2.1 with respect to clinical decision-making (see chapter 8).

Finally, sound policy-making must be based on reliable and valid data. Increasingly, transparency and accountability are required in the framework of the evidence-based approach. We therefore have to know what the epidemiological data actually add to the existing body of knowledge and whether they really reflect the extent of the presence of the clinically relevant psychiatric disorders in the general population. Since a number of surveys generate quite different prevalence rates, the results of the CIDI and the SCAN, these being the major representatives of the two psychiatric interview traditions, were directly compared with one another (see chapter 9).

4.4 Research questions

This thesis uses descriptive epidemiological data generated by the SCAN-2.1 by way of focusing on the relevance of clinical judgement in psychiatric diagnoses, the objective being to improve the quality of the diagnostic process and consequently of the quality of epidemiological data.

Part II: Descriptive epidemiology, i.e. case finding: the prevalence and distribution of identified cases

In chapter 5 (monitoring) the primary question that we address is whether the prevalence rate of neurotic and functional psychotic disorders and their distribution changed over the period 1983 to 1997. In this chapter, we set out to answer the following, more specific questions:

1. What were the prevalence and distribution of psychiatric disorders in 1983 and 1997?
2. What were the shifts in psychiatric prevalence and its socio-demographic distribution between 1983 and 1997?
3. If there was a shift, how could it be explained in terms of time trends (age, cohort and period effects)?

In chapter 6 the NHA-2 project of 1997 is described as an example of mapping. In this chapter, we attempted to answer these questions:

1. What were the prevalence rate and the distribution of the total range of psychiatric monomorbid and comorbid disorders in 1997?
2. What is the additional value of the three diagnostic categories, which can be regarded as an expansion of the diagnostic range determined in the general population with DSM-IV en SCAN-2.1?

Part III: Diagnostics: case identification

In chapter 7 the psychometric properties of the SCAN-2.1 are described. The specific questions requiring an answer are:

1. Can the semi-structured, clinically oriented SCAN-2.1 interview be reliably applied in a stratified sample of the open population?
2. What is the validity of the SCAN-2.1 diagnoses, given the standard score of experienced clinical psychiatrists?

To make clinical judgement operational, it is important to recognize the aspects of signs and symptoms that should be clinically assessed in the process of case identification. In chapter 8, the following questions about clinical decision-making in SCAN-2.1 are answered:

1. Does clinical judgement, as operationalised in the SCAN-2.1, make a difference?
2. What is the effect of the scores for the sub-clinical items (illness but not disease) upon the prevalence rates?

In order to improve our understanding of the common finding that fully structured interviews in the general population consistently give rise to discrepant findings in comparison with clinically oriented psychiatric diagnostic interviews, in chapter 9 we report the outcome of a direct comparison of the results of the SCAN with the results obtained with the CIDI. We set out to answer the following questions:

1. In what respects do results obtained with the SCAN-2.1 and the CIDI agree and differ with reference to the same subjects, chosen from the open population?
2. What is the influence of clinical judgement in this regard?

Part IV: Discussion and conclusions

In chapter 10 we summarise and discuss the epidemiological outcomes and the results with respect to the use of clinical judgement in interview instruments. Thereafter we conclude with some recommendations.

Part II

5 Monitoring: Psychiatric disorders in a Dutch health area – a repeated cross-sectional survey with the PSE

5.1 Introduction

Time and again it has been suggested that the prevalence rate of psychiatric disorders is rising owing to the noxious factors that characterise the way of life in modern society. Many aspects of contemporary life, such as increasing urbanism, mounting bureaucratisation and governmental regulation, and the rapidity of social change in the spheres of family life and work, are viewed as creating psychological difficulties for people. There is also evidence to the contrary. For example, in spite of the numerous social changes in Stirling County between 1952 and 1970, Murphy et al. (1984) found a stable point prevalence rate of depression and anxiety disorders. Likewise, Nandi et al. (2000) reported an unchanged overall level of psychiatric morbidity over twenty years in a rural Indian community, but an interesting alteration of the morbidity pattern: the rates of anxiety, phobia and hysteria had fallen dramatically and those of depression and mania had risen significantly. In London, Bebbington et al. (1981; 1997) even found slightly lower rates than those they had determined in the 1980s. On the other hand, there was a rise in 1-year psychiatric prevalence rate from 20% to 28% between the two waves of the Epidemiological Catchment Area (ECA) cohort study in the USA (Regier et al. 1998). The diagnostic validity of the latter results has been questioned as they were generated by lay interviewers administering fully structured diagnostic instruments instead of by experienced clinicians using a semi-structured interview (Cooper & Singh 2000). Some feel that self-report interviews capture a wider range of trivial or less severe symptoms and disorders than a psychiatrist would, and that the high estimates of prevalence rates are biased (Eaton et al. 2000; Frances 1998).

Studies on the question of whether psychiatric prevalence is declining, stable, or increasing are fraught with methodological difficulties. While a prospective cohort study has the obvious advantage of consistency as far as the population, the diagnostic system, and the measuring instruments are concerned, secular population changes cannot be examined adequately that way, because the cohort ages over time. A repeated cross-sectional survey, on the other hand, does offer that possibility, but with the disadvantage of a relative incomparability of the study populations, the diagnostic systems, and the measuring instruments. Because of these difficulties, there are few studies that compare community psychiatric surveys carried out in the same area at different points in time.

Despite the widely differing reports of psychiatric prevalence rates, a more or less consistent relationship has been found between psychiatric disorder and sociodemographic characteristics (Dohrenwend & Dohrenwend 1974; Surtees et al. 1983). In their review, the last-named authors reported a higher prevalence rate for females, urban people, the elderly, the poorly educated, the unemployed, and divorced and widowed persons. This finding may reflect a higher degree of exposure and/or vulnerability to the stresses of everyday life. Consistent with Rothman's definitions on sufficient and component cause (see chapter 2 on epidemiology), vulnerability can be described as the existence of a (group of) component cause(s) that increase the risk of a disease manifestation but that by itself is not sufficient to do so. Vulnerability increases the risk that another component cause, normally not powerful enough to elicit the disease, will do so. This is called the eliciting factor in the vulnerability-stress model, the equivalent of the precipitating factor of the aetiological approach.

At the end of the 1970s a research project was started to determine any change in psychiatric prevalence rates over a 14-year period, and to test the hypothesis of a consistent association between demographic characteristics and psychopathology. This was done by means of a repeated cross-sectional survey in the Nijmegen Health Area (NHA) in the Netherlands, in which a semi-structured psychiatric instrument administered by clinically experienced interviewers was used.
In a previous paper, Hodiamont et al. (1987) described a community study of psychiatric disorders among 18 to 64 year olds that was performed 1983. The investigators used the General Health Questionnaire (GHQ) and the ninth version of the Present State Examination (PSE-9), which were combined in a two-phase design, the object having been to estimate the prevalence rate of psychiatric disorders. The relationship between the PSE-9 caseness and the GHQ score was expressed in a logistic regression model that yielded a 7.3% (± 1.9%)[ii] point-prevalence of the PSE cases. There was no significant difference between males and females. Higher case rates were found in the age range of 55-59 years, among divorced and widowed persons, for lower educational and occupational levels, the unemployed, the chronically ill, the work-disabled and city dwellers. The study added further evidence to a growing body of epidemiological data, suggesting similar figures and patterns of psychiatric disorder in populations in industrialised countries (Bebbington et al. 1981; Henderson et al. 1979; Lehtinen et al. 1990; Vazquez-Barquero 1990). However, it raised new questions about age-specific risks and the magnitude of the difference in case rates between urban and rural inhabitants of this rather densely populated region (Hodiamont et al. 1992).
The objectives of the 1997 survey were **to compare the case rates** and their **distribution** in 1983 and 1997, to explore **the impact of urbanisation** on these

findings, and **to explore possible time trends in terms of cohort, period, and age effects** on psychiatric prevalence. Because the foremost concern was to assure similar research conditions, basically the same design was used as in 1983.

5.2 Methods

5.2.1 The sample

In 1983, a random sample of 4,500 persons was drawn from the population of all adults in the NHA (over 250,000 people, aged 18-64). Persons who had been admitted to institutions or hospitals or who did not have a sufficient command of the Dutch language were not interviewed. The area was divided into two parts: Nijmegen, and the remaining 29 communities. A total of 10 communities were selected from the second stratum without replacement. They were drawn one at a time and the chance of selection was proportional to the size of the community. Each of these 10 communities supplied a systematic sample of 280 people and Nijmegen supplied 1,700 people. In this way an equal proportion of individuals from Nijmegen and the other communities in the sample was obtained. The distribution by gender, age and marital status of the 3,245 subjects in phase 1 (T1) corresponded to the population distribution according to the 1982 census data (CBS).

In 1997, again, a random sample of 4,517 persons was drawn from the population of all non-institutionalised Dutch-speaking adults in the NHA (over 325,000, aged 18-74; nearly 295,000 aged 18-64), which consisted of the city of Nijmegen and the surrounding communities (SC). In order to gain easy access to medical data, 32 general practitioners (GPs) were recruited, 11 working in the city of Nijmegen and 21 in the surrounding communities. They all met the following criteria: they had worked in the same practice within the NHA for at least 2 years, and they had been supported by an operational automated information system for patient data for at least 1 year. Since virtually every inhabitant of the Netherlands is registered in a general practice, the degree of this registration is equivalent to that of the registry offices (Boerma et al. 1993).

An age- and gender-stratified sample of 150 persons from each practice were sent a letter by their GP asking for their informed consent to take part in the survey. The total sample of 4,517 persons agreed well with reference to the population distribution for age and gender according to the 1997 CBS data. After two rounds of reminders, 2,049 persons consented, 1,975 explicitly refused, and no answer was received from 493. An interview was conducted with 1,813 persons in T1, 1,617 of whom were aged 18-64 years (the survey sample reported on in this chapter).

To establish whether a bias had affected the response, data on gender, age, degree of urbanisation and the prescription of psychotropic medication (as a parameter of mental illness) were used to test whether the survey sample was an adequate representation of the general population. Census data on other sociodemographic and economic characteristics are not regularly registered in the Netherlands.

No significant difference in GP prescription of psychotropic medication was found between the survey sample and the general population, so that selection bias with respect to psychiatric disorders was unlikely. Females and rural dwellers were overrepresented in both survey samples (1983 and 1997), and the elderly only in 1997 (Table 5.2.1). Adjustments for these differences were made in the statistical analyses. Like the 1983 results, the 1997 results may therefore be considered representative of the general population.

5.2.2 Procedure

Because the clinical instruments that were available at the time, i.e. the PSE-9 (Wing 1974) and version 2.1 of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN-2.1) (Wing et al. 1990), are extensive and expensive interviews, a two-phase strategy was adopted (Duncan-Jones 1979; Henderson et al. 1979; Williams et al. 1980) both in 1983 and in 1997, in which the 30-item General Health Questionnaire (GHQ) (Goldberg 1972; Goldberg 1978; Koeter & Ormel 1991b) was used to screen in T1, and the PSE-9 and the SCAN-2.1 were administered respectively to determine the existence of psychiatric cases and to specify these cases in phase 2 (T2).

In 1983 one out of eight subjects was randomly selected from the T1 survey sample and invited to enter T2, while in 1997 one out of four subjects was randomly selected from the T1 survey sample. In both cases, predominantly those persons with a low score on the GHQ-30 were selected. The two random sub-samples were supplemented in 1983 by the T1 subjects who had a GHQ-30 score of at least 10 and in 1997 by the T1 subjects who had a GHQ-30 score of at least 4. This was done in order to get sufficient data for the estimation of the logistic model and, subsequently, the psychiatric prevalence rate.

This approach led, in 1983, to 3,232 completed GHQs and 486 PSE-9 interviews. In 1997, 1,605 subjects completed the GHQ and 707 were interviewed with the SCAN-2.1.

5.2.3 The general interview

The T1 interview with the survey sample covered sociodemographic characteristics, and contained questionnaires on social support, coping behaviour, life events, chronic disease, psychiatric history, quality of life, political and religious convictions, and also the GHQ.

Table 5.2.1

Comparison of the data from the 1983 and 1997 survey samples (NHA-1 and NHA-2) with CBS data for degree of urbanisation, gender, and age, and representativeness of the 1997 survey sample for psychotropic medication prescription with reference to GP practices.

		1983			1997		
		Survey sample N = 3,245 %	Population N = 247,141 %	χ^2	Survey sample N =1,617 %	Population N = 294,782 %	χ^2
Urbanisation	Nijmegen	33.6	39.1	p < .001	29.2	34.3	p < .001
	Surrounding communities	66.4	60.9		70.8	65.7	
Gender	Males	48.7	50.5	p = .037	44.3	50.4	p < .001
	Females	51.3	49.5		55.7	49.6	
Age (years)	18-19	5.2	5.9	p < .001	1.5	3.6	p < .001
	20-24	14.7	15.3		5.9	11.1	
	25-29	15.5	14.7		9.2	13.2	
	30-34	14.9	14.0		13.2	13.3	
	35-39	12.7	11.5		14.0	12.5	
	40-44	10.3	9.3		11.5	11.8	
	45-49	6.9	8.5		9.3	11.7	
	50-54	7.6	7.8		8.5	8.7	
	55-59	7.4	7.1		7.8	7.5	
	60-64	4.9	5.9		7.1	6.7	

Prescriptions	Survey sample N = 1,617	All patients N = 80,315	
Antidepressants	3.4	3.4	.98 (NS)
Benzodiazepines	10.3	9.5	.21 (NS)
Antipsychotics	0.4	0.8	.07 (NS)
Analgesics	10.3	9.6	.33 (NS)
Other psychotropic drugs	2.7	2.7	.88 (NS)

Consistent with our NHA-1 report (Hodiamont et al. 1987), the discussion on the comparison is confined to gender, age, marital status, highest completed education level, employment status, and degree of urbanisation.

For the statistical analyses, the degree of urbanisation was reduced to two categories: Nijmegen (large city) and the SC in the health area.

5.2.4 The measurement of psychiatric disorder

The SCAN-2.1 is a semi-structured, clinical psychiatric interview. It includes the tenth version of the Present State Examination (PSE-10), the successor to the PSE-9, developed by the WHO (Wing 1996). As in the NHA-1, a one-month reference period was chosen for the current psychiatric state.

Because of their common roots and design, it is possible to convert the PSE-10 scores into the PSE-9 scores. In co-operation with Fokko Nienhuis of the WHO-SCAN Collaborating Centre in Groningen, an automated program for generating the PSE-9 output from the SCAN 2.1 data was developed. Of the 140 PSE-9 items, 97 were identical to the items of the PSE-10, so that the item scores could be converted linearly. For example, zero and subliminal ratings of depressed mood in the PSE-10 were converted into a zero score in the PSE-9, whereas clinically significant (2) and severe (3) ratings were converted into scores of 1 and 2, respectively, for the PSE-9. A total of 37 composite PSE-9 items were broken up into single PSE-10 items. For instance, the scores of the nine items on specific phobia in the PSE-10 covering the sole composite item of the PSE-9 on this pathology were converted into the latter on the basis of their clinical significance. Only six[iii] of the 140 PSE-9 items were not integrated into PSE-10, three of which were not included in CATEGO-4, the algorithm program for the PSE-9, and one of which was irrelevant to the diagnostic process, and consequently their scores could not be converted. The two remaining items were not included in the PSE-10 as they yielded no relevant information when the PSE-9 was administered. After the conversion, the same CATEGO-4 program that was used in 1983 was used to classify the 1997 data in terms of the Index of Definition (ID, with eight levels based on the presence of discriminating symptoms, key combinations of symptoms and the total number of symptoms present) and tentative psychiatric diagnoses based on the ICD-9. A tentative psychiatric diagnosis was calculated at or above ID level 5 (threshold or caseness level).

In view of the clinical nature of the PSE and the SCAN, physicians, psychologists, and psychiatric nurses were selected for their psychiatric experience as interviewers. All interviewers participated in a 1-week training course given by Professor Frans Verhey Sr. (Maastricht) in 1983 and Professor Rob Giel and Fokko Nienhuis of the WHO-SCAN collaborating centre (Groningen) in 1997. Booster sessions were given to enhance their expertise. The reliability of the PSE-9 and the SCAN-2.1 turned out to be satisfactory (Duine et al. 1985a; Rijnders et al. 2000).

5.2.5 Analyses

The SAS version 8 package was used for all analyses.

The relation between the GHQ score and the PSE caseness was expressed in a logistic regression model. In this model, the logit of an individual's caseness probability is expressed as a linear function of the GHQ score:

$$\log p/1\text{-}p = \alpha + (\beta \times \text{GHQ score}),$$

where p is the probability that a certain GHQ score corresponds to a PSE case and α and β are parameters estimated from the outcome of the T2 interviews (Henderson et al. 1979; Hodiamont & Veling 1984).

All sociodemographic variables were tested to determine whether they had a significant impact on the logistic regression model. Since only the two degrees of urbanisation resulted in significantly different models, different parameters for Nijmegen and the surrounding communities were used. The 1983 data were re-analysed in this respect.

The prevalence rate of the PSE cases was estimated on the basis of the regression equation and the distribution of GHQ scores from the first phase of the study in which the sample had participated. To estimate the 95% two-sided confidence intervals, the logistic model was integrated into a bootstrap procedure, a re-sampling technique with duplication (Dunn et al. 1999; Efron & Tibshirani 1993). On the assumption of a normal distribution of the estimated prevalence rate, it was tested whether any two subgroups differed in respect of case rates. Because of the nominal character of the variables, in this way the case rates of each pair of subgroups was tested for significance with a Kruskal Wallis test.

The prevalence rates of the ID levels and of the tentative diagnoses were calculated in two steps. In the first step, the prevalence rate was estimated separately for two strata of the PSE subjects in T2: for those subjects with GHQ scores of < 10 and for those with GHQ scores of ≥ 10. In the second step, the mean of the two prevalence rates was calculated by weighing on the basis of the proportional size of the two strata as they occurred in T1. For the differences between the survey sample and the general population with respect to gender, degree of urbanisation, and age, adjustments were made in the statistical analyses.

5.3 Results

5.3.1 Sociodemographic shifts

The NHA population, apart from having aged markedly, grew in 14 years by almost 20%, mainly (91%) on account of newly built areas in the small towns and villages of the SC, while the population of Nijmegen hardly changed.
From the data (e.g., Table 5.2.1 and Table 5.3.5), it is clear that the sociodemographic profile of the population altered markedly after 1983. With regard to marital status, for example, an increase in the number of single, divorced and cohabiting persons was observed. The number of persons remunerated for their work grew substantially, unemployment fell, and the percentage of full-time housewives was halved. On the other hand, the percentage of chronically ill or work-disabled persons remained generally unaltered. The mean highest completed level of education increased, which resulted in a smaller contribution by persons who had only a primary school education. Overall, the contribution of persons to the lowest levels of the various socioeconomic classes appeared to have decreased.

5.3.2 GHQ data

Table 5.3.1 shows the distributions of GHQ-30 scores by gender, degree of urbanisation (Nijmegen and SC), and year of survey (the NHA-1, 1983; the NHA-2, 1997). In both surveys, females scored higher than males and Nijmegen subjects scored higher than those of the SC. The mean scores show a significant increase over time, most markedly for females in the city of Nijmegen (+2.7, $p < 0.001$). While the differences between Nijmegen and the SC were minimal in 1983, in 1997 the score for Nijmegen was significantly higher than the mean score for the SC.
The increase concerned especially the items on social dysfunction. Items on social interaction in particular were scored worse (for instance, ‘finding it easy to get on with other people’; ‘able to feel warmth and affection for those close to you’). On the other hand, items on socio-economic prospects (for instance, ‘hopeful about your own future’; ‘finding that life is a constant struggle’) were scored unchanged or even better.
In short, although the socio-economic status of the population improved, judging by the aforementioned shift from 1983 to 1997, its health status in terms of GHQ scores worsened.

Table 5.3.1
Frequency distribution* of GHQ-30 scores by gender, degree of urbanisation and year of survey

	Males				Females			
	Nijmegen		SC		Nijmegen		SC	
	1983	1997	1983	1997	1983	1997	1983	1997
	N = 512	N = 191	N=1060	N = 519	N = 571	N = 276	N=1089	N = 619
GHQ-score	%	%	%	%	%	%	%	%
0 – 2	62.6	54.2	64.7	63.2	59.9	45.2	63.3	55.8
3 – 5	18.4	18.1	17.5	14.8	18.5	18.8	18.1	14.9
6 – 7	5.7	5.0	5.9	6.1	7.0	6.1	6.1	5.0
8 – 9	2.5	3.8	4.4	4.1	4.9	6.1	4.1	6.3
≥ 10	10.8	18.9	7.4	11.8	9.8	23.9	8.4	18.1
Mean score	3.2	5.2	2.9	3.6	3.4	6.1	3.0	4.7
S.D.†	2.1	3.8	1.9	2.9	2.1	4.0	2.0	3.5

* Adjusted for age
† Standard deviation

5.3.3 Relationship between GHQ and the PSE

On the basis of the T2 interviews, maximum-likelihood estimates were obtained for the parameters in the logistic regression equation, which applied to all of the demographic variables except degree of urbanisation (see Figure 5.3.1).
The relationship between the GHQ and the PSE manifested itself in quite different curves for Nijmegen (N'n) and the SC respectively in 1983, whereas the shape and course of both curves looked quite alike in 1997.
The higher GHQ scores are associated with a higher probability of caseness in 1983 than in 1997, while the lower GHQ scores show a higher probability of caseness in 1997 than in 1983.

Probability of PSE-caseness by GHQ-score in Nijmegen (N'n) and the SC, 1983 and 1997

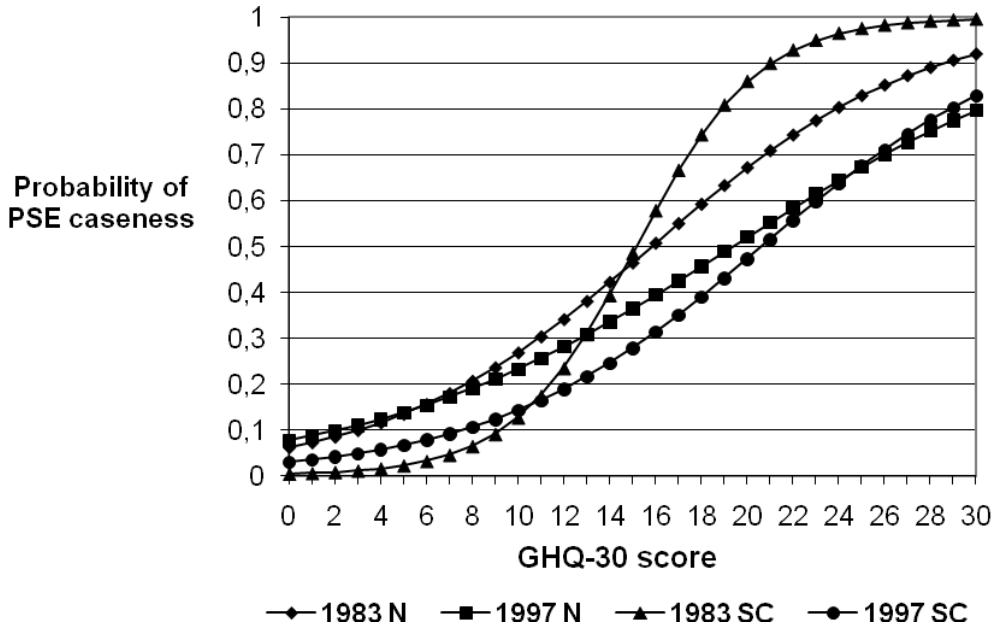

Figure 5.3.1

5.3.4 PSE data

ID level

The percentages of the various PSE-ID levels in the NHA-1 and the NHA-2 are presented in Table 5.3.2.

With respect to 1983, the number of persons with symptoms at or above threshold level (ID≥5) increased. Contrary to our expectations, this also applied to the number of persons *without* psychiatric symptoms (ID level = 1). Thus, the 1997 ID levels were percentually higher for the extremes than was the case in 1983 ($p = 0.010$). This held even after adjustment for the significant effect ($p = 0.003$) of degree of urbanisation.

The prevalence rates of CATEGO classes and corresponding tentative ICD diagnoses presented in Table 5.3.3 show a two to five times higher level in Nijmegen than in the SC ($p = 0.011$). The prevalence rate of psychotic states found in the NHA-2 was more than twice that in the NHA-1 for both Nijmegen and the SC, whereas for neurotic states, a difference was found for the SC only. In Nijmegen, the prevalence rate of neurotic states tended to be stable over time.

Table 5.3.2
Prevalence* of PSE-ID levels in the NHA-1 and the NHA-2 by degree of urbanisation

		Nijmegen		SC	
		1983	1997	1983	1997
		N = 1083	N = 464	N = 2149	N=1138
ID-level		%	%	%	%
1	No PSE symptoms	25.6 ± 2.4 †	47.6 ± 7.4	33.1 ± 2.7	47.6± 6.0
2 3 4	Symptoms below threshold level	59.8 ± 0.6	34.9 ± 3.5	62.2 ± 0.6	43.7 ± 1.9
5	Symptoms at threshold level	9.6 ± 1.2	10.3 ± 0.5	3.8 ± 2.2	5.2 ± 1.8
6 7 8	Symptoms above threshold level	5.0 ± 1.8	7.3 ± 3.4	1.0 ± 1.1	3.5 ± 2.4

* Adjusted for gender and age
† 95%= confidence interval

PSE-caseness

The overall prevalence rate of PSE cases in the population of the NHA as a whole was 11.9% (±2.7%)[iv] in 1997, significantly higher than the prevalence rate of 7.8% (±2.3%), which was found in 1983 ($p < 0.001$). The prevalence rate of depression and anxiety disorders for the health area as a whole, was also significantly different between1997 (9.7% ±2.9%) and 1983 (7.7% ±2.2%, Table 5.3.3; $p < 0.01$).

Table 5.3.3
Prevalence* of PSE-CATEGO classes (with the ICD equivalents) in the NHA-1 and the NHA-2 by degree of urbanisation

Description	CATEGO classes	ICD code	Nijmegen 1983 N = 1083 %	Nijmegen 1997 N = 464 %	SC 1983 N = 2149 %	SC 1997 N = 1138 %
Schizophrenic psychoses	S	295.3	1.7 ± 0.6 †	3.7 ± 0.7	0.3 ± 0.3	0.7 ± 0.2
Paranoid states	P	297.9				
Manic and mixed affective psychoses	M	296.1,3				
Depressive psychoses	D	296.2	12.7 ± 2.2	13.9 ± 3.2	4.4 ± 3.0	7.5 ± 3.9
Inhibited depressions	R	296.2 or 300.4				
Neurotic depressions	N	300.4				
Anxiety neuroses	A	300.0,2				

* Adjusted for age and gender
† 95%= confidence interval

5.3.5 The relation between PSE data and sociodemographic variables

Urbanisation

The resulting prevalence rates for degree of urbanisation and gender are presented in Table 5.3.4.

The increase in case rates turned out to be smaller for Nijmegen (from 12.8 ±5.3 to 18.0 ±6.4) than for the SC (from 4.5 ±1.4 to 8.8 ±2.5). Although there was still a substantial difference in psychiatric prevalence rate between city and rural dwellers, there was a tendency for the surrounding communities to catch up in terms of psychiatric caseness. In terms of diagnoses, there was a two-fold increase in the prevalence rate of psychotic disorders, both in Nijmegen and in the SC. For neurotic disorders, however, a more or less stable situation was found in Nijmegen, in contrast to a near doubling in the SC.

Table 5.3.4
Prevalence rates in the NHA-1 and the NHA-2 by degree of urbanisation and gender

	1983	1997
Nijmegen	12.8 ± 5.3 †	18.0 ± 6.3
SC	4.5 ± 1.4	8.8 ± 2.5
Males	7.5 ± 2.2	10.8 ± 2.7
Females	8.0 ± 2.3	13.1 ± 2.8

† 95%= confidence interval

Gender

In contrast to the similar PSE case rates for males (7.5 ± 2.2%) and females (7.8 ± 2.3) in 1983, the rates in 1997 differed significantly for males (10.8 ± 2.7%) and females (13.1 ± 2.8%, $p<0.001$).

Age

Our conclusion in 1983 that 'for both sexes the case prevalence rate more or less increases with age up to about age 60' should be readjusted. Table shows a higher rate for persons aged 45 years and older, but no clear pattern for the younger age groups. In contrast to our findings in 1983, when most cases were found in the 55- to 64-year-old age group, caseness in 1997 peaked in the 45- to 54-year-old age group for both sexes, most strikingly in the SC. In addition, a lower case rate was found for females aged 55-64 from the SC in 1997 than in 1983.

Marital status

The 1983 finding of a significantly higher case rate for divorced females than for females married or living with others was confirmed in 1997. The tendency for single males to be at high risk from psychiatric caseness turned out to be consistent over time in Nijmegen, and also applied to single males in the SC in 1997.

Employment status

A significant difference in the rates of disorder was found when the unemployed, chronically ill and work-disabled were compared with all other categories, irrespective of gender, degree of urbanisation, and year of inquiry ($p < 0.05$). The only deviation from the general increase in case rates for all employment statuses over time (most markedly for chronically ill, male, urban dwellers; $p = 0.024$) was the significant decrease in chronically ill, male, rural dwellers ($p = 0 .018$).

Level of education

There was a trend towards an inverse relationship between case rate and education level, which reached significance in the SC ($p < 0.05$). For female city dwellers with the lowest levels of education, the rise in the prevalence rate of the PSE caseness from 1983 to 1997 was significantly greater than for most other categories.

Table 5.3.5
Sociodemographic distribution of PSE caseness (Males)

		Males							
		Nijmegen (N'n)				SC			
		1983		1997		1983		1997	
		N = 512		N = 191		N = 1060		N = 519	
		N	%	N	%	N	%	N	%
Age group	18-24	115	12.9	16	18.9	198	3.4 [b]	28	5.6 [b*]
	25-34	158	10.3	44	16.1	338	3.1 [b d]	93	6.2 [b d]
	35-44	108	13.8	64	15.9	254	4.7	151	7.8
	45-54	62	12.1	41	18.0	156	5.3 [a]	144	10.2 [a]
	55-64	69	16.2	26	17.7	114	6.9 [c]	103	8.0 [c]
Marital status	Single	65	14.4	42	19.2	21	2.4	28	10.2
	Living with parents	58	10.9	5	15.9	206	3.2 [b]	32	6.8 [b]
	Living with partner	73	12.6	38	14.6	54	6.4	35	9.0
	Married	296	12.2	93	16.6	760	4.3 [b]	401	7.4 [b]
	Divorced	15	13.9	12	14.1	14	13.1 [a]	22	11.4 [a]
	Widowed	5	29.6	1	12.4	5	3.2	1	3.5
Employment status	Full-time employed	259	10.3 [b d]	132	14.4 [bd]	746	2.8 [f]	419	7.2 [f]
	Part-time employed	31	9.8 [b]	23	13.6 [b]	42	1.0 [f]	25	7.8 [f]
	Retired	20	10.6 [b]	3	15.4 [b]	11	1.2 [f]	14	10.3 [f]
	Chronically ill/ unable to work	42	22.3 [a]	5	49.4 [a *]	79	19.3 [e]	33	12.9 [e *]
	Unemployed	68	16.8 [c]	11	30.2 [c]	110	6.5 [f]	12	11.9 [f]
	School or college student	87	13.2 [b]	12	20.9 [b]	70	1.6 [f]	12	7.3 [f]
	Householder	1	6.1	--	--	--	--	1	3.0
	Others without an occupation	4	8.5	5	22.4	2	25.0	2	11.4
Level of education	University	68	9.6	43	16.8	69	1.1	36	8.5
	Higher vocational education	97	9.9	33	15.5	189	2.5 [b]	95	7.3 [b]
	Upper-stream sec. education	64	14.0	35	17.0	56	2.6	36	5.0
	Lower-stream sec. & post-sec. lower vocational education	49	12.0	23	15.0	142	4.2	124	7.9
	Lower-stream sec. education	34	14.3	14	17.2	67	3.8	51	8.2
	Continued primary education	119	13.4	33	18.6	348	4.4	137	8.6
	No education or only primary education	81	15.8	10	19.0	189	7.8 [a]	40	10.9 [a]

In some subgroups, due to missing data the cumulative N was lower than the N of the survey sample stated in the title of the table.

Age group
SC males: Kruskal-Wallis .01 < p < .05; Scheffé (.05) [a - b] and [c - d]
* significant difference between the year pairs (p < .05)
Marital status
SC males: Kruskal-Wallis .001 < p < .01; Scheffé (.05) [a - b]
Employment status
N'n males: Kruskal-Wallis p < .001; Scheffé (.05) [a - b] and [c - d]
SC males: Kruskal-Wallis p < .001; Scheffé (.05) [e - f]
Level of education
SC males: Kruskal-Wallis .01 < p < .05; Scheffé (.05) [a - b]
* significant difference between the year pairs (p < .05)

Table 5.3.5 (continued)
Sociodemographic distribution of PSE caseness (Females)

		Females							
		Nijmegen (N'n)				SC			
		1983		1997		1983		1997	
		N = 571		N = 276		N = 1089		N = 619	
		N	%	N	%	N	%	N	%
Age group	18-24	135	12.2	29	16.9	194	3.4 [f]	47	6.4 [f]
	25-34	146	12.0	86	18.2	339	3.8	136	11.6 [*]
	35-44	113	12.6	80	17.3	267	4.4	166	8.8
	45-54	87	14.0	53	21.6	165	5.6 [e]	154	13.0 [e]
	55-64	90	15.6	28	22.8	124	8.4	116	7.4 [*]
Marital status	Single	72	11.0 [d]	49	16.7 [d]	29	6.4	16	4.6
	Living with parents	42	13.3	3	20.8	120	2.8 [f h]	24	9.4 [f]
	Living with partner	86	11.5 [d]	93	17.4 [d]	41	4.9	49	8.7
	Married	322	12.9 [d]	106	19.8 [d]	839	4.3 [f]	495	10.4 [f]
	Divorced	30	22.0 [c]	22	26.0 [c]	27	13.5 [e]	18	13.5 [e]
	Widowed	19	16.5	2	20.9	33	13.2 [g]	17	8.9
Employment status	Full-time employed	87	11.9 [h]	101	14.4 [h]	187	2.6 [j]	102	10.0 [j *]
	Part-time employed	118	11.2	94	18.6	211	4.4	258	9.3
	Retired	2	7.2	2	8.3	1	0.5	6	5.6
	Chronically ill/ unable to work	8	17.7 [g]	7	31.7 [g]	13	5.9 [i]	12	23.0 [I]
	Unemployed	24	12.8	10	32.3	29	9.7	9	16.9
	School or college student	72	12.6	15	16.7	45	3.8 [j]	18	7.0 [j]
	Householder	257	14.2	34	26.6 [*]	602	5.2 [j]	190	10.3 [j]
	Others without an occupation	3	23.7	12	16.3	1	0.3	25	9.6
Level of education	University	33	14.3	68	15.2	27	3.4	9	9.2
	Higher vocational education	101	10.2	82	14.6	136	1.6 [d]	96	7.5 [d]
	Upper-stream sec. education	65	11.7	35	17.0	57	2.9	55	11.3
	Lower-stream sec. & post-sec. lower vocational education	74	13.1	12	24.2	167	3.4	169	8.2
	Lower-stream sec. education	57	12.0	23	19.5	122	2.2	113	11.2 [*]
	Continued primary education	104	13.7	40	25.8 [*]	322	5.4	143	11.8
	No education or only primary education	137	15.4	16	32.0 [*]	258	7.9 [c]	33	12.2 [c]

In some subgroups, due to missing data the cumulative N was lower than the N of the survey sample stated in the title of the table.

Age group
SC females: Kruskal-Wallis .01 < p < .05; Scheffé (.05) [e - f]
* significant difference between the year pairs (p < .05)

Marital status
N'n females: Kruskal-Wallis p < .001; Scheffé (.05) [c - d]
SC females: Kruskal-Wallis .001 < p < .01; Scheffé (.05) [e - f] and [g - h]

Employment status
N'n females: Kruskal-Wallis p < .001; Scheffé (.05) [g - h]
SC females: Kruskal-Wallis .001 < p < .01; Scheffé (.05) [i - j]
* significant difference between the year pairs (p <. 05)

Level of education
SC females: Kruskal-Wallis .001< p < .01; Scheffé (.05) [c - d]
* significant difference between the year pairs (p < .05)

5.4 Discussion

As stated in the introduction, the objectives of this monitoring study were **to compare the case rates** and their **distributions** in 1983 and 1997, to identify **the influence of urbanisation** on the changes in prevalence rates and distribution, and **to explore possible cohort, period, and age effects** on the prevalence rate of PSE cases.

5.4.1 Study limitations

The NHA studies used a cross-sectional design and this had an advantage and a disadvantage: the advantage was the possibility of examining secular population changes, and the disadvantage concerned potential limitations with regard to the comparability of the study populations and the compatibility of the measuring instruments and the diagnostic system used.

At first sight, the notable non-response in 1997 appears to call into question the assumption of a representative survey sample and limit the comparability of the results for 1983 and 1997. Potential subjects seemed to be deterred by the exhaustive information campaign deemed necessary for informed consent by the medical ethics committee. On the other hand, this sampling design, in which GP practices were used, enabled us to check for selectivity by means of psychotropic medication. There was no evidence of a selection bias in the sense of patients with psychiatric morbidity (medicated) being overrepresented in the sample. We therefore assume that for 1997 too, the sample was a fair representation of the general population.

So that limitations would be reduced with regard to differences between interview instruments and diagnostic systems used in 1983 and 1997, the NHA-2 instrument, i.e. the SCAN-2.1/PSE-10, was deliberately chosen on the basis of its suitability as the successor of the the PSE-9 used in the NHA-1, which made possible a WHO-approved conversion of the NHA-2 scores into the NHA-1 format. In this way the data from the NHA-1 and the NHA-2 could be compared properly.

5.4.2 Prevalence rates, GHQ – PSE relationship

Table 5.3.1 shows that the number of persons with a GHQ-score $\geq$ 10 underwent a two-fold increase from 1983 to 1997. Since Figure 5.3.1 shows a lower associated rate of caseness for the higher GHQ-scores for 1997 than for 1983, it can be concluded that in 1997 more subjects reported a substantial feeling of decreased well-being without an underlying formal psychiatric disorder than in 1983.

In the 14-year interval between the two studies, newly built areas in the small towns and villages of the SC were partially populated as a result of migration from the city of Nijmegen. Such a population shift offers an explanation of the relative similarity of the course and shape of the curves representing the GHQ – PSE relationship for Nijmegen and the SC in 1997 in comparison with 1983.

5.4.3 Prevalence rates, the PSE-caseness

The increase in the prevalence rates of the PSE-ID levels 5 through 8 from 1983 to 1997 could have been expected from the distribution of the GHQ-scores, and is consistent with the view that after conversion the SCAN-2.1 adequately detects PSE-cases. For the lower PSE-ID score range however, in 1997 the prevalence rate of level 1 was higher, while levels 2 through 4 were lower than in 1983. Although the SCAN is sensitive to PSE-9 caseness, its scoring algorithm might also help to explain this discrepancy. In contrast to the PSE-9, the SCAN can be used to score signs and symptoms that are present, but at a sub-clinical level (score 1). In the more common neurotic sections this feature might result in the absence of mild symptoms when the SCAN scores are converted to PSE-9 scores.
Through logistic regression, the one-month prevalence rate of neurotic and functional psychotic caseness, which was rated by means of a clinical semi-structured psychiatric interview after screening with the aid of a questionnaire, was estimated at 7.8 $\pm$ 2.3% (after reanalysis) in 1983 and 11.9 $\pm$ 2.7% in 1997, in representative samples of a Dutch general population, aged 18-64 years. Consequently, the prevalence rate of psychiatric disorder had risen by about 50% for the NHA as a whole. The mathematical explanation of this phenomenon is as follows:
On the one hand, in 1997, a low GHQ-score was associated with a higher probability of caseness than in 1983 (see Figure 5.3.1). This concerned the majority of the 1997 subjects (Table 5.2.1). On the other hand, despite the fact that for high GHQ-scores the probability of caseness slightly waned in 1997 in comparison with 1983, this subgroup had its boosting effect on caseness since its size more than doubled over the years (Table 5.3.1).

To put the prevalence rates in perspective, they should be compared with rates from other surveys conducted analogously. In the 1980s, the Dutch PSE-9 case rates (ID $\geq$ 5: 7.8%) (Hodiamont et al. 1987) were compared and contrasted with those for Canberra (ID $\geq$ 5: 9.1%) (Henderson et al. 1979) and Camberwell (ID $\geq$ 5: 10.9%) (Bebbington et al. 1981), and the Dutch rates were found to be the lowest. This finding still applies in comparison with prevalence estimates from other surveys done in the 1980s using the same instruments in Finland and Spain (Vazquez-Barquero 1990). Our estimate of 11.9% for 1997 fits in

better with the overall pattern of psychiatric disorder in western industrialised countries. Whereas our prevalence rate increased by more than 50% in 14 years, Bebbington et al. (1981; 1997), using more or less the same design and instruments as those used in the NHA study, reported a more or less stable one-month prevalence rate over about the same period (10.9% and 9.8%, respectively). Kessler et al. (2005c) recently reported a stable psychiatric 12-month prevalence rate over the period 1990 - 2003 (29.4% and 30.5%, respectively). These investigators compared data from the NCS study with findings from the NCS replication study. Nandi et al. (2000) reported the same effect on a rural Indian area from 1972 to 1992 (11.7% and 10.5%, respectively), and Merikangas et al. (2003) did so from their Zurich Cohort Study, conducted over a period of 20 years. They reported the prevalence rate of depressive and anxiety disorders to be 4.9%, 8.6%, 9.7%, and 7.7% in 1981, 1986, 1988, and 1993 respectively. The latter investigation had the advantage of being prospective, and the disadvantage of being mainly based on a narrow age range, namely 19-20 year old in 1979.

5.4.4 Sociodemographic shift and distribution of psychiatric disorders

The sociodemographic profile of the population changed markedly after 1983. The NHA population, apart from having aged markedly, grew by almost 20% in 14 years, mainly on account of the SC. Furthermore, there was an increase in divorced persons and singles, and a decrease in persons living with their parents. Thirdly, the mean highest completed level of education increased. Finally, as the percentage of persons remunerated for their work grew substantially, the percentage of full-time housewives was halved, and the percentage of chronically ill or unemployed persons decreased for males, but remained generally unaltered for females. When looking for explanations of the case-rate findings, It is important to reflect on these (major) changes at the societal level since societal interventions are potentially most powerful from the preventive point of view (Susser et al. 2002).
Over time, our psychiatric case rates increased in all sociodemographic categories, but the increases were greater in the lower social classes. This latter finding may be an artefact created by the decreasing number of people in these classes, but there are other explanations.

Despite the widely differing reports of psychiatric prevalence rates, a more or less consistent relationship has been found between psychiatric disorder and sociodemographic variables (Dohrenwend & Dohrenwend 1974; Surtees et al. 1983). In their reviews, the last-named investigators reported a higher prevalence rate for females, urban dwellers, the elderly, the poorly educated,

the unemployed, and divorced and widowed persons (Fryers et al. 2003; Fryers et al. 2005).
In the NHA-1 (1983), the assumption of a higher percentage of disorders among city dwellers, those who had lost a partner, poorly educated people, and the unemployed was confirmed. Only the gender difference was not substantiated. On the basis of the NHA-2 (1997) data, the aforementioned starting hypothesis was reconfirmed in all respects, including the higher case rate for females. Consequently, the former explanation that “role patterns in the Nijmegen Health Area may have changed in such a way that the inherent stress produced a similar amount of psychiatric disorder for the two sexes” (Hodiamont et al. 1987) can no longer be considered tenable.
From 1983 to 1997 the prevalence rates for females increased to a greater extent than for males. From a societal point of view, this might be explained by the growing demand for well-educated employees in Western society. This factor has reinforced the assumption, generally accepted today, that all individuals should become part of the professional work force, which has stimulated females to catch up with males in respect of the highest education levels, this pressure having been significantly lower in 1983. In parallel, the work status of females changed from running a household to paid part-time employment. Simultaneously, an increasing percentage of females decided to cohabit (with or without children in the living unit) instead of living with their parents or on their own. Altogether these changes were tantamount to an increase in role stress, which constitutes a risk factor with reference to the burden side in the vulnerability-stress model. To a lesser degree the same dynamic applied to males, who took on increasing responsibilities on the domestic front in addition to their existing professional duties.
Apart from role stress, possible explanations that have been suggested for the apparent association between gender and the prevalence rate of mental disorders are response bias and biological vulnerability. Gender did not have any effect on the GHQ-PSE model, so that response bias is an unlikely explanation for our data. If biological and/or hormonal vulnerability does indeed play a role in the higher rate of disorders among females, one might expect a reduction of this difference between males and females for the category of people over 55 years of age in comparison with younger persons. So far, there has been some evidence to this effect: for all persons over 55 years of age the prevalence rate is between 10.4% and 10.9%, while for persons 54 years and younger, females score higher than males.

5.4.5 Influences of urbanisation

As for the apparent influence of degree of urbanisation on the rate of psychiatric disorders in 1983 and 1997, it can be inferred from the parameters of the different logistic regression models that, for lower GHQ scores, people from the city of Nijmegen had a higher probability of PSE caseness than those from the SC. In the course of time, however, this difference tended to decline (see Figure 5.3.1). This trend is especially true for the **neurotic states**, for which the SC subjects effectively caught up with the Nijmegen subjects through almost doubling the prevalence rate over the 14-year interval, whereas the prevalence rates for **psychosis** increased in both regions.

A societal explanation might be that the newcomers moved into newly built areas, outside the old village centres with their sense of strong social coherence. This change would have increased the feeling of anonymity, resulting in a strongly diminished sense of social cohesion in comparison with their previous city life. The decline of social coherence might have resulted in an increased vulnerability (McConnell et al. 2002), more distinct in the SC than in Nijmegen.

Another possible explanation is the difference in the sociodemographic shift between 1983 and 1997 in Nijmegen and in the SC. In the latter, people tended to a greater extent than city dwellers to remain locked into the lower socio-economic classes (poor education, non-remunerated work, divorced or widowed). Kirkbride et al., in their study on psychotic syndromes (2006), concluded that "the incidence of all diagnoses was greater in Southeast London than Nottingham or Bristol after standardisation for age and gender. These differences remained after further adjustment for ethnicity, except for affective disorders. This suggests truly 'psychotogenic' effects of that environment or population stratification in terms of psychosis risk."

Altogether, these societal changes are likely to have contributed to an influence of urbanisation on the overall increase in the prevalence rate from 1983 to 1997.

5.4.6 Cohort, period, and age effects

The high rates of disorder found in 1983 in the group aged 55-59 years persisted even after controls for marital status, employment status, and perceived morbidity had been introduced. This fact prompted Hodiamont et al. (1987) to speculate at the time that the experience of having lived in a war zone had had an effect on his cohort (a so-called cohort effect); after all, his population sample had been in their teens during the Second World War. If the war had played any role in the above-mentioned high rates, one would expect

to find a peak prevalence in the age range of 69-73 years in 1997, but this proved not to be the case.
The increase in prevalence rate between 1983 and 1997 may have been caused by one or more, possibly cumulative, **period effects**. A first **period effect** was reported in the paper by Hodiamont at al. (2005), in which we explained that there had been an overall increase (from 19.0% in 1983 up to 28.9% in 1997) in the GHQ-30 scores for Nijmegen and the SC over the 14-year interval in question. This increase especially concerned the items on social dysfunction. Items on social interaction in particular were scored worse. On the other hand, items on socio-economic prospects were scored unchanged or even better. In short, although the socio-economic status of the population in the area as a whole improved, on the basis of the aforementioned shift in social interaction from 1983 to 1997, its health status in terms of GHQ scores actually worsened. These findings are supported by Verhaak et al. (2005), who reported on the same issues in a large-scale, national, Dutch replication study based on GP practices. Among other results, these researchers observed a national trend towards a substantial increase in mental health problems over a 14-year interval (1987 – 2001), the GHQ-12 having been used as a measuring instrument. In this study the GHQ score ≥ 2 rose from 16.8% in 1987 up to 22.8% in 2001, with females, people with a lower education level and widowed and divorced persons having significantly higher scores. These subjects presented an increase in family and relational problems, while material problems were less often reported.
A second **period effect** for the area as a whole had to do with the above-mentioned remarkable sociodemographic shift in the NHA over the relatively short 14-year interval (see 6.4.4). Verhaak et al. (2005) reported the same phenomenon nation-wide (See Verhaak Table 1). It may therefore be concluded that this shift is not just a coincidental regional finding. The assumption is that this upward sociodemographic shift results in changed social roles, with attendant increasing role stress, and increasing prevalence rates as a consequence. A third general **period effect** may be found in the policy of reducing the number of psychiatric hospital beds, resulting in a 12% decrease in admission days (Ypsilon 2001), under the influence of which the prevalence rates for **psychosis** increased in both Nijmegen and the SC. In contrast to the preceding, the prevalence rate of **neurotic** disorders in the SC nearly doubled in contrast to a more or less stable prevalence rate of these disorders in Nijmegen, which contributed to the higher overall prevalence rate for the NHA as a whole in 1997 (Table 5.3.3). A fourth (but regionally oriented) **period effect** is the aforementioned migration of city dwellers to the SC, who brought their **neurotic** burden in the form of a higher prevalence rate with them. A fifth explanation is a (geography-specific) **period effect**: the evacuation of many of the rural dwellers because of the high waters of the Rhine and the impending

inundation of their homes in the winters of 1993 and 1995. Some support for this hypothesis may be found in the fact that the prevalence rate of neurotic disorders in the communities adjoining the river was nearly 25% higher ($p = .07$) than in the inland communities.

In this monitoring study, a clear **age effect** could not be proved. An age effect implies, after all, that the highest case rate should be consistently found in the same age group. Because most cases were found in the age range of 45-54 years in 1997 (in contrast to 1983, when such cases massed in persons aged 55-64 years), this condition was not borne out in a strict sense. However, the possibility that the expected age effect was obscured by the above-mentioned period effects can not be ruled out. The 45-54 yr age group may have been more prone to the various period effects, and may have reacted more morbidly in terms of neurotic symptoms than the resigned seniors.

5.4.7 Conclusions

All things considered, it can be concluded that, while the distribution of cases remained generally unaltered, the prevalence rate of psychiatric disorders rose substantially in the Nijmegen Health Area from 1983 to 1997. In sum, neither an age effect nor a cohort effect was found. However, striking period effects showed up. These concerned both the population as a whole – apparently because of overall changes in social roles and a modified health policy – and the rural communities, as a result of the sociodemographic shift due to the above-noted migration and of the imminent inundation of their homes and property.

Despite the improved socio-economic conditions in the overall population surveyed, the increasing complexity of life, which has already been referred to, apparently took its toll, even among those best equipped to deal with the socio-affective vicissitudes. This conclusion seems to contradict the general finding that upward mobility in society is associated with a better health status. This finding, however, concerns upward-moving individuals in rather steady social classes. In this study, an improvement in socio-economic conditions was found for the great majority of the population. An undesirable result of this socio-economic improvement might be a destabilisation of social classes, manifesting itself in feelings of indistinctness and uncertainty among the members of these classes and a greater overall vulnerability to stress.

These social changes, which Murphy et al. (1984) hypothesised were potentially harmful for mental health, were definitely extensive in the Nijmegen Health Area (Hodiamont et al. 2005) and the Netherlands as a whole (Verhaak et al. 2005), and probably were limited in other countries. Murphy et al. stated that the impersonalness and anonymity associated with increasing urbanism might be

destroying the resources that people use in order to achieve a sense of inner worth, that mounting bureaucratisation might be stifling autonomy and independence, and that the rapid pace of change in family life and work might be taxing the individual's ability to adapt and maintain personality integration. In the same vein, Stevens (1999) pointed out the likelihood of a connection between various forms of psychopathology and Western society's inability to satisfy the archetypal needs of the human species. He quoted John Bowlby, who had declared that "the further the rearing environment deviates from the environment of evolutionary adaptedness, the greater the likelihood of pathological development". However, the question of whether psychiatric prevalence in general is rising, stable, or decreasing in Western countries, and if so, under what conditions, has not been answered as yet.

References

Bebbington PE, Hurry J, Tennant C, Sturt E and Wing JK. (1981) Epidemiology of mental disorders in Camberwell. *Psychological Medicine,* 11, 561-579.

Bebbington PE, Marsden L and Brewin CR. (1997) The need for psychiatric treatment in the general population: the Camberwell Needs for Care Survey. *Psychological Medicine,* 27, 821-834.

Boersma WGW, de Jong FAJM and Mulder PH. (1993) The Netherlands. In *Health Care and General Practice across Europe*. pp. 77-80. Nivel (Netherlands Institute of Primary Health Care), Utrecht.

Cooper B and Singh B. (2000) Population research and mental health policy. Bridging the gap [editorial]. *British Journal of Psychiatry,* 176, 407-411.

Dohrenwend BP and Dohrenwend BS. (1974) Social and cultural influences on psychopathology. *Annual Review of Psychology,* 25, 417-452.

Duine TJ, Hodiamont PPG, Peer PGM and Boerma LH. (1985) Betrouwbaarheidsaspecten van de Present State Examination in vier verschillende populaties [in Dutch]. *Tijdschrift voor Psychiatrie,* 27, 326-340.

Duncan-Jones P. (1979) Validity and uses of the GHQ [letter]. *British Journal of Psychiatry,* 135, 382.

Dunn G, Pickles A, Tansella M and Vazquez-Barquero JL. (1999) Two-phase epidemiological surveys in psychiatric research [editorial]. *British Journal of Psychiatry,* 174, 95-100.

Eaton WW, Neufeld K, Chen LS and Cai G. (2000) A comparison of self-report and clinical diagnostic interviews for depression: diagnostic interview schedule and schedules for clinical assessment in neuropsychiatry in the Baltimore epidemiologic catchment area follow-up. *Archives of General Psychiatry,* 57, 217-222.

Efron B and Tibshirani RJ. (1993) *Introduction to the bootstrap*. (Vols. 57) Chapman and Hall, New York.

Frances A. (1998) Problems of defining clinical significance in epidemiologiccal studies. *Archives of General Psychiatry,* 55, 119.

Fryers T, Melzer D and Jenkins R. (2003) Social inequalities and the common mental disorders: a systematic review of the evidence. *Social Psychiatry and Psychiatric Epidemiology,* 38, 229-237.

Fryers T, Melzer D, Jenkins R and Brugha T. (2005) The distribution of the common mental disorders: social inequalities in Europe. *Clinical practice and epidemiology in mental health,* 1, 1-12.

Goldberg DP. (1972) *The detection of psychiatric illness by questionnaire*. Oxford University Press, London.

Goldberg DP. (1978) *Manual of the General Health Questionnaire*. NFER Publishing Company, Windsor.

Henderson S, Duncan-Jones P, Byrne DG, Scott R and Adcock S. (1979) Psychiatric disorder in Canberra. A standardised study of prevalence. *Acta Psychiatrica Scandinavica,* 60, 355-374.

Hodiamont PPG, Peer N and Syben N. (1987) Epidemiological aspects of psychiatric disorder in a Dutch health area. *Psychological Medicine,* 17, 495-505.

Hodiamont PPG, Rijnders CATh, Mulder J and Furer JW. (2005) Psychiatric disorders in a Dutch Health Area: a repeated cross-sectional survey. *Journal of Affective Disorders,* 84, 77-83.

Hodiamont PPG, Syben N, Koeter MW and Oldehinkel AJ. (1992) The relation of psychiatric disorder and urbanisation [in Dutch]. *Tijdschrift voor Psychiatrie,* 34, 426-438.

Hodiamont PPG and Veling SHJ. (1984) A model to assess psychiatric prevalence: The relation GHQ-PSE. [in Dutch]. *Tijdschrift voor Psychiatrie,* 26, 592-601.

Kessler RC, Demler O, Frank RG, Olfson M, Pincus HA, Walters EE, Wang P, Wells KB and Zaslavsky AM. (2005) Prevalence and treatment of mental disorders, 1990 to 2003. *New England Journal of Medicine,* 352, 2515-2523.

Kirkbride JB, Fearon P, Morgan C, Dazzan P, Morgan K, Tarrant J, Lloyd T, Holloway J, Hutchinson G, Leff JP, Mallett RM, Harrison GL, Murray RM and Jones PB. (2006) Heterogeneity in incidence rates of schizophrenia and other psychotic syndromes: findings from the 3-center AeSOP study. *Archives of General Psychiatry,* 63, 250-258.

Koeter MW and Ormel J. (1991) *General Health Questionnaire: manual [in Dutch]*. Swets and Zeitlinger, Lisse.

Lehtinen V, Lindholm T, Veyola J and Väisänen E. (1990) The prevalence of PSE-Catego disorders in a Finish adult population. *Social Psychiatry and Psychiatric Epidemiology,* 25, 187-192.

McConnell P, Bebbington P, McClelland R, Gillespie K and Houghton S. (2002) Prevalence of psychiatric disorder and the need for psychiatric care in Northern Ireland. Population study in the District of Derry. *British Journal of Psychiatry,* 181, 214-219.

Merikangas KR, Zhang H, Avenevoli S, Acharyya S, Neuenschwander M and Angst J. (2003) Longitudinal trajectories of depression and anxiety in a prospective community study: the Zurich Cohort Study. *Archives of General Psychiatry,* 60, 993-1000.

Murphy JM, Sobol AM, Neff RK, Olivier DC and Leighton AH. (1984) Stability of prevalence. *Archives of General Psychiatry,* 41, 990-997.

Nandi DN, Banerjee G, Mukherjee SP, Ghosg A, Nandi PS and Nandi S. (2000) Psychiatric morbidity of a rural Indian community. *British Journal of Psychiatry,* 176, 351-356.

Regier DA, Kaelber CT, Rae DS, Farmer ME, Knauper B, Kessler RC and Norquist GS. (1998) Limitations of diagnostic criteria and assessment instruments for mental disorders. Implications for research and policy. *Archives of General Psychiatry,* 55, 109-115.

Rijnders CATh, van den Berg JF, Hodiamont PP, Nienhuis FJ, Furer JW, Mulder J and Giel R. (2000) Psychometric properties of the schedules for clinical assessment in neuropsychiatry (SCAN-2.1). *Social Psychiatry and Psychiatric Epidemiology,* 35, 348-352.

Stevens A. (1999) Evolutionary psychiatry: a new perspective on the explanation and treatment of mental disorders. In *Routes in the GGZ*. pp. 45-51. Tilburg University Press, Tilburg.

Surtees PG, Dean C, Ingham JG, Kreitman NB, Miller PM and Sashidharan SP. (1983) Psychiatric disorder in women from an Edinburgh community: associations with demographic factors. *British Journal of Psychiatry,* 142, 238-246.

Susser E, Bresnahan M and Link B. (2002) Peering into the future of psychiatric epidemiology. In MT Tsuang en M Tohen (Eds.), *Textbook in psychiatric epidemiology, second edition*. pp. 195-211. Wiley-Liss, New York.

Vazquez-Barquero JL. (1990) Mental health in primary-care settings. In D Goldberg en D Tantam (Eds.), *The public health impact of mental disorder*. pp. 35-44. Hogrefe & Huber Publishers, Toronto.

Verhaak PF, Hoeymans N, Garssen AA and Westert GP. (2005) Mental health in the Dutch population and in general practice: 1987-2001. *British Journal of General Practice,* 55, 770-775.

Williams P, Tarnopolsky A and Hand D. (1980) Case definition and case identification in psychiatric epidemiology: review and assessment. *Psychological Medicine,* 10, 101-114.

Wing JK. (1974) *Measurement and classification of psychiatric symptoms*. Cambridge University Press, Cambridge.

Wing JK. (1996) SCAN and the PSE tradition. *Social Psychiatry and Psychiatric Epidemiology,* 31, 50-54.

Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D and Sartorius N. (1990) SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry,* 47, 589-593.

Ypsilon (2001). Survey: Reduction of hospital bed does not result in 'Italian rumpus' [in Dutch]. *Schizofrenie Bulletin*.

6 Mapping: The Nijmegen Health Area-2 project

6.1 Introduction

Descriptive epidemiology, and mapping in particular, are basic means of turning up leads through which patterns in the occurrence of disease can be disclosed (Hennekens & Burning 1987). Mapping provides essential information about the *who, where* and *when* of the disorders in question. It improves our understanding of the relationship between the disorder and biological, socio-demographic, psychological and social characteristics. As a result, the aetiology of the pathological process becomes clearer (Lilienfeld & Lilienfeld 1992). Mapping is therefore an indispensable first step in the search for causality (Susser et al. 2002). This process is especially important as it is generally accepted in medicine that causal treatment, if available, is the best option.

From the moment that Quality-Adjusted Life Years (QALYs) became a primary measure of disease burden, psychiatric disorders were ranked among the top five health problems of the Western World. Depression and anxiety disorders affect a sizeable portion of the population and often result in a substantial loss of quality of life for a prolonged period of time. Consequently, reducing the number of QALYs lost on account of psychiatric disorders is one of the main targets of health policy. Given the continuous need of health policy makers for information on how to achieve this goal most effectively, “the relevance of mapping psychiatric prevalence in the general population is undisputed” (Thornicroft & Tansella 2001).

Our scores, when converted into the PSE-9 format, suggested a significant increase in psychiatric prevalence rates over time (see chapter 5 of this thesis and Hodiamont et al. [2005]). Findings of increased psychiatric prevalence rates are frequently criticised. Cooper & Singh (2000) reported that, apart from an increase in the prevalence rates of psychiatric disorders in the population over the previous 10 – 15 years, rates of psychiatric morbidity tended to inflate. This inflation might be due to the use of different time frames (lifetime instead of 1-month prevalence rates), the employment of ‘non-clinical’, fully standardised instruments instead of clinical, semi-structured interview instruments (Brugha et al. 1999a), and the actual evolution of interview instruments, mostly resulting in a more extensive and a more sensitive successor. Another explanation for the rise in prevalence rates is the evolution of classification systems, which entails an expanded diagnostic range or more diagnostic subtypes (Batstra et al. 2002).

As the primary objectiveof the Nijmegen Health Area-2 (NHA-2) survey was to obtain data in the NHA for comparison with the NHA-1 data (see chapter 5), it was decided to use the 1997 state-of-the-art clinical instrument, the successor to the PSE-9, namely Schedules for Clinical Assessment in Neuropsychiatry

(SCAN-2.1), to map the psychiatric point (1-month) prevalence rate. This choice resulted in a number of qualitative advantages.

First, the use of the SCAN-2.1 implied an assessment from the **clinical viewpoint**, whereas most other population studies employed a fully (and strictly) standardised instrument, which makes interviews conducted by lay interviewers suitable, for example the Composite International Diagnostic Interview (CIDI). To arrive at a valid clinical judgement, the SCAN interviewers must probe until they are satisfied with the information obtained, since it is the professional who must decide if symptom definitions are fulfilled. As a consequence, the subjective symptoms reported are weighted in the SCAN-2.1 for their clinical relevance, possibly resulting in more moderate prevalence rates than those found in studies with lay interviewers (Brugha et al. 1999a; Brugha et al. 2001; Cooper & Singh 2000).

Second, the SCAN-2.1 is the first (clinical) interview instrument that uses the **integral DSM-IV classification system** as a diagnostic reference (disorders usually first diagnosed in childhood being an exception). Use of the integral diagnostic reference creates the only opportunity to classify all disorders, which makes possible, on the one hand, a comparison with other recent field studies conducted with other instruments, and, on the other hand, the use of a more integral classification system with respect to the previous clinical interviews, which were based on a more limited diagnostic reference. the SCAN has been used before (Andrews et al. 1995; Brugha et al. 2001; Eaton et al. 2000), but never in a general population survey with the integral DSM-IV range. The NHA study is the first mapping in the general population to have adopted the full DSM-IV range of psychiatric disorders diagnosed on the basis of a direct clinical judgement. This approach is rarely used. Reasons for the rarity of use were given by Narrow et al. (2002), who stated: *'In large epidemiologic surveys, direct clinical judgement is rarely used because of the high cost of clinical time and the large number of subjects, so proxy measures are used.'* The use of the full DSM-IV range might lead to an increase in prevalence rates – indeed, on reflection, the real rates – in comparison with studies in which a partial range is used, which gives rise to artificially low prevalence rates.

Third, the SCAN-2.1 is an **extensive** interview instrument, with over 1800 questions, in comparison with the 140 questions of the PSE-9. Thanks to this comprehensive questioning, the SCAN-2.1 has the advantage of generating detailed, clinically relevant data, making possible precise classification on a subdiagnostic level. As a result, comorbidity can be expected to increase because of the more detailed classification of disorders. As the SCAN-2.1 is a semi-structured instrument that allows flexibility in accordance with the respondents' actual situation, the duration of the interview (an average of 90 minutes, with peaks of over 3 hours) appears not to bother the respondents, and it generates a reliable picture of the clinical psychiatric situation.

Finally, because the SCAN-2.1 is the first psychiatric diagnostic instrument into which the full diagnostic range of DSM-IV is incorporated, the opportunity is created to pay attention to the expansion of the full adult diagnostic range. Compared with the PSE-9, this means an **extention of the diagnostic range** to include diagnostic categories such as substance-related, somatoform, sleep, and dissociative disorders. Compared with the CIDI, which assesses the prevalence of almost all psychiatric disorders, the diagnostic extension concentrates on sleep disorders. In the international literature on epidemiologic studies in the general population, somatoform, dissociative, and sleep disorders are hardly, if ever, reported. In this study, in addition to the prevalence rate worked out on the basis of the full diagnostic range, we shall concentrate on three 'newly' determined categories in the general population (somatoform, sleep, and dissociative disorders) by identifying (1) the contribution of the separate prevalence rates for the three new categories to the overall prevalence rates; (2) their supplementary contribution to comorbidity; and (3) their contribution to possible changes in the stable pattern of prevalence distributions (Dohrenwend & Dohrenwend 1974).

As argued before, any qualitative change in the diagnostic process has its influence on the quantitative diagnostic outcome, and consequently on the rate of the co-existence of diagnoses, i.e. comorbidity. Comorbidity itself is a major health issue, because, on the one hand, it points to the complexity of the underlying problems, and on the other, most interventions (medication, psychotherapy, etc.) are not comorbidity 'proof', in that they are focussed on the treatment of well-defined monomorbid disorders. Many definitions of comorbidity are at hand, each contributing to a wide range of prevalence rates, ranging from less than 1% to as high as 50% (van den Akker et al. 1996). The recession of the hierarchical approach in the transition from DSM-II to DSM-III and its successors has inevitably led to an increase in the prevalence rates of comorbid psychiatric disorders (Kessler et al. 1997; Pincus et al. 2004).
In this study we shall use the definition of comorbidity adopted by Feinstein (1970): '*any distinct additional entity that has existed or may occur during the clinical course of a patient who has the index disease under study'.* However, we shall restrict comorbidity to formal psychiatric disorders in order to examine its prevalence rates and sociodemographic distribution in the general population, and to study the impact of the three newly determined categories on both of these phenomena.

At the same time as the NHA-2 study was performed, the first wave of the NEMESIS study, a national survey on psychiatric morbidity, was carried out (Bijl et al. 1998) with a non-clinical, fully standardised, instrument (the CIDI-1.1) based on a narrower diagnostic range and derived from the DSM-III-R. The

opportunity to compare *mutatis mutandis* the non-clinical NEMESIS results with the clinical NHA-2 findings was an important reason for the Ministry of Health, Welfare, and Sport to support the NHA-2 study.
In the Netherlands, the SCAN had never been used in a general population survey. Internationally, only a few studies are known in which the SCAN-1.0 had been used (McConnell et al. 2002; Roca et al. 1999).

In this chapter, mapping is illustrated through reporting the epidemiological prevalence data and their distribution as found in the NHA-2 survey, carried out in 1997 with the SCAN-2.1 as a clinical interview instrument for the first time in a general population survey. The objective of the NHA-2 study, in addition to the comparison with the NHA-1 (see chapter 5), was to assess the following items with respect to the **integral diagnostic reference** of the DSM-IV:

1. the **prevalence** rates of psychiatric disorders in the general population;
2. the sociodemographic **distribution** of these disorders;
3. the rate of **comorbidity** and its distribution with reference to gender, age and urbanisation;

to study

4. the influence of **three newly determined diagnostic categories** on prevalence, distribution and comorbidity;

and to compare

5. the NHA-2 data, with data from a contemporary epidemiological national survey in the Netherlands (**NEMESIS**) with respect to prevalence rates, distribution, and comorbidity.

6.2 Methods

6.2.1 History of the NHA-2

In the early 1990s, Hodiamont, who had been responsible for the psychiatric part of the NHA-1, was approached by a representative of the Ministry of Health, Welfare and Sport, who suggested a replication of the psychiatric part of the NHA-1 project. Given that the primary objectiveof the 1997 survey (NHA-2) was to obtain (replication) data in the Nijmegen Health Area (NHA) for comparison with the 1983 NHA-1 data, the opportunity was seized to use an up-to-date interview instrument with optimal possibilities for conversion of the scores to the 1983 format. The comparison of the NHA-1 and NHA-2 data with the use of the PSE-9 format was described in chapter 5.

In this chapter, the 1997 data are presented with the aid of the data obtained with the interview instrument (SCAN-2.1) that was actually used, in order to describe the prevalence rates and distribution of psychiatric disorders in the general population.

In addition to being used for the objectives mentioned above in this chapter and those stated in chapter 5, the data of the NHA-2 study were employed in six other studies. The purpose of one of these was to develop a list of High Risk Indicators (HRI) for psychiatric caseness for use in primary care practice (Roscam Abbing & Hodiamont 2000). A second study aimed at generating a screening list in primary care practice for the risk of future long-term use of benzodiazepines (Zandstra et al. 2002a; Zandstra et al. 2002b; Zandstra et al. 2004). The object of the third study was to determine the relationships between social factors and depressive symptoms in the general population (Meertens et al. 2003; Meertens 2004). In the fourth study, Kooij et al. (2005) assessed the internal and external validity of attention-deficit hyperactivity disorder in a population-based sample of adults. The fifth study explored psychiatric comorbidity with a view to gaining a better insight into its magnitude, its distribution and its consequences for the general population and also into the differences between monomorbidity and comorbidity (Rijnders & Furer 2003). Finally, in the sixth study, which was explorative, the diagnosis and treatment of mental health problems in general practice were investigated (van Rijswijk et al. 2000a; van Rijswijk et al. 2000b; van Rijswijk et al. 2005; van Rijswijk 2005).

6.2.2 Study design

Directly in line with the design of the NHA-1 study (1983), the NHA-2 study consisted of a two-stage (sampling), two-phase (interview) procedure (Duncan-Jones 1979; Henderson et al. 1979; Hodiamont et al. 1987; Hodiamont & Veling 1984). As far as the assessment of prevalence is concerned, the differences

between the first and the second study included an extension of the research domain and a refinement of the research method. In particular, the NHA-1 domain was extended, which implied a widening of the diagnostic spectrum to include, for example, substance-related, somatoform, dissociative, and sleep disorders, and an increase in the age range up to 75 years. The refinement concerned mainly the screening instruments.

6.2.3 Survey sample

In order to gain easy access to the medical data needed for the additional projects, it was decided to use general practice registers for stage one. Since practically every inhabitant of the Netherlands is registered in a general practice, the degree of this registration is comparable to that of the registry offices (Boerma et al. 1993). We recruited 32 general practitioners (GPs), 11 working in the city of Nijmegen and 21 in the surrounding communities. They all fulfilled the following criteria: they had worked in the same practice within the Nijmegen Health Area (NHA) for at least 2 years and had been supported by an operational, automated information system for patient data for at least the year preceding the study.

The 32 GP registers contained information on 80,315 persons aged 18 to 75 years, and a data set was constructed for all these persons, containing gender, age and prescription of psychotropic drugs. A random sample of 4,517 persons consisting of an equal number of subjects from each practice were sent a letter by their GP asking if they would give informed consent to take part in the survey (stage two). Those who consented were interviewed over a period extending from September 1997 to March 1998.

6.2.4 Interviews

The interviews were conducted in accordance with a two-phase design with the General Health Questionnaire 30 items (GHQ-30) as a screening list for psychiatric disturbance in phase one (T1) and the SCAN-2.1 psychiatric interview in phase two (T2).

After informed consent was obtained from the respondents at stage two of the sampling, in T1, experienced interviewers from the Institute of Applied Sociology held a general interview with all subjects in the survey sample. In addition to the GHQ, which was used for screening in T1 (Furer et al. 1995a; Koeter & Ormel 1991a; Ormel et al. 1989), the interview included questionnaires about health measurement and health-related concepts[v] and also addressed demographic characteristics. After completion of T1, the GPs, with the help of their automated system, filled out a questionnaire on the following: consultation frequency, GP psychiatric diagnosis and treatment (if

applicable), comorbidity data, some physical morbidity data, and the reason for the use of benzodiazepines.
In T2, the respondents were interviewed with the SCAN-2.1, which includes the latest (tenth) version of the PSE (PSE-10), the successor to the PSE-9 (Wing et al. 1990; Wing 1996). The PSE-10 consists of two parts, the first dealing with anxiety, mood, and other neurotic manifestations and the second with psychotic and cognitive phenomena, and observation items on speech, behaviour, and affect. In this survey, the interview with the July 1997 version of the SCAN-2.1 was used (Giel & Nienhuis 1997).
Because the SCAN-2.1 requires clinically experienced trained interviewers and is time-consuming and consequently expensive, not all respondents in T1 were interviewed in T2. The number of respondents entering T2 was limited to a randomly assigned 25% portion of the survey sample, predominantly with a low score on the GHQ-30. To amplify for psychiatric caseness, the random sample was supplemented with all respondents who had a score of 4 or higher on the GHQ-30 in T1. The rationale for this selection was based, on the one hand, on the expectation that we would detect a sufficient number of cases to enter T2, and, on the other hand, on the requirement to include respondents with a low score, which is essential for analysing the full-scale GHQ-SCAN relationship.
Consistent with the clinical, semi-structured nature of the SCAN-2.1, 30 psychologists and psychiatric nurses were selected for their knowledge of psychopathology. Special sessions on psychiatric phenomenology were organised to optimize their actual clinical judgement. All interviewers participated in a 1-week training course on the use of the pencil-and-paper version of the SCAN, organised by the WHO-SCAN collaborating centre in Groningen. To assure data of good psychometric quality (see chapter 7), regular booster sessions with the interviewers were organised to improve their knowledge of psychopathology and interviewing techniques. Furthermore, a WHO-approved, computerised version of the SCAN-2.1 was introduced, including an algorithm for calculating DSM-IV diagnoses. During the fieldwork, problems encountered during the data checks were discussed with a psychiatrist (CR). Before the survey was conducted, a reliability study was carried out. The claim that clinical interviews are imperfect because of low interrater reliability and potential validity problems (Wittchen et al. 1999) has been rebutted for the PSE-9 (Duine et al. 1985b; Lesage et al. 1991; Wing et al. 1977) and for the SCAN-2.1 (Andrews et al. 1995; Brugha et al. 1999b; Cheng et al. 2001; Rijnders et al. 2000; Tomov & Nikolov 1990).

6.2.5 Analyses

For all analyses, the Statistical Analyses for Social Sciences software package (SAS v8.2) was used.

For all SCAN data, DSM-IV diagnoses were computed by means of the SCAN-2.1 algorithm. For the purpose of this survey, SCAN diagnoses were grouped on three levels: first, closely related diagnoses like single episode and recurrent depressive disorder were sorted into one *diagnostic group* (e.g., depressive disorder); then, associated diagnostic groups (e.g., depressive, dysthymic, and bipolar disorders) were classified into one *diagnostic category* (e.g., mood disorder), in accordance with the structure of DSM-IV; finally, the highest level of aggregation was represented by *diagnostic caseness,* defined as "the presence of at least one specific diagnosis".

The relationship between the GHQ and SCAN "caseness" was expressed in a logistic regression model in which the probability of an individual's qualifying for "caseness" was formulated as a linear function of the GHQ score:

$$p(\text{caseness}) = \frac{1}{1+e^{-(\alpha + \beta * GHQ)}}$$

where p indicates the probability that a certain GHQ score corresponds to SCAN-caseness, on the basis of the assessed α and β,which reflect the relationship between GHQ and the SCAN caseness as estimated from the outcome of all T2 interviews (Henderson et al. 1979; Hodiamont & Veling 1984). Because only 25% of the respondents with a GHQ score of less than 4 during their T1 interview were interviewed in T2, a weighting procedure was used for this group to restore the proportions initially present in the T1 group. Through substitution of the assessed probability for SCAN caseness for the GHQ score for all respondents in T1, the prevalence rate for the population was calculated.

The prevalence rates for the diagnostic groups and categories were calculated through weighing the extent to which they occurred in the two GHQ classes with scores < 4 and ≥ 4 for T2 respondents. The weighting factors were determined by the ratio of the sizes of the two GHQ classes to one another in the first phase of the sample.
So that the data for the general population could be interpreted, a post-stratification weighting procedure was applied with the aid of 16 defined strata for gender, age (4), and urbanisation (2) on the basis of the 1997 census data (CBS 1997).
To estimate the 95% two-sided confidence intervals for these weighted data, the logistic model was integrated into a bootstrap procedure, i.e. a re-sampling technique with duplication (Dunn et al. 1999; Efron & Tibshirani 1993).
The significance of the difference in prevalence rates within sociodemographic distributions was tested in an ANOVA model.

6.3 Results

6.3.1 Response and representativeness

The response rate of the survey sample (T1) was 40.3% (N = 1,813). A total of 440 subjects in the survey sample were randomly assigned to T2 (subgroup 1), supplemented by 509 subjects with a GHQ-30 score ≥ 4 (subgroup 2). Finally, 767 subjects were interviewed in T2, the response rate being 81% in both subgroups.

The impact of non-response was analysed with respect to the representativeness of general practices (stage 1 of the survey), sociodemographic variables, and psychotropic drug use as a measure of mental health (respondent-aligned, stage 2 of the survey).
The general practices were representative of all Dutch general practices with regard to practice size, number of GPs working in the practice, GPs' gender and year of registration, total working time of GPs and the percentage of practices with a pharmacy. The only significant difference was the higher number of practices with training facilities for GPs among the participating practices (χ^2 = 5.6, p = .018). Since being a GP trainer was related to neither caseness nor GHQ score, it was not necessary to correct for this finding.
Gender, age, and urbanisation (the city of Nijmegen versus the smaller surrounding communities) were used to test whether the survey sample was a good reflection of the general population (CBS 1997). Both the GP population and the GP sample proved to be good representations of the regional general population (Table 6.3.1). The significances found were a consequence of the larger numbers and are not relevant percentually.
In the survey sample in T1, males and inhabitants of Nijmegen had a lower representation and younger persons were underrepresented. For all three sociodemographic variables, a correction was made for the analysis, in which the 1997 CBS census data were used.
The existence of at least one prescription of psychotropic medication over the 12-month period preceding the study was used as an indicator of mental illness. In epidemiological research into psychiatric prevalence, this can be regarded as an important representativeness indicator. No differences were found between the GP population and the samples with respect to any of the five medication categories (Table 6.3.2).

Table 6.3.1
Representativeness for gender, urbanisation, and age

Sociodemographic variable		NHA population N=325,566 %	GP population N = 80,315 %	χ^2	(p)	GP sample N = 4,517 %	χ^2	(p)	Survey sample N = 1,813 %	χ^2	(p)
Gender	Male	49.9	49.3	17.6		49.6	0.18		44.7	19.8	
	Female	50.1	50.7	(<.001)		51.4	(NS)		55.3	(< .001)	
Urbanisation	Nijmegen	34.5	35.0	13.6		34.7	0.05		29.3	21.6	
	Surrounding communities	65.5	65.0	(<.001)		65.3	(NS)		70.7	(<.001)	
Age (yrs)	18 - 29	25.1	25.7			24.5			14.9		
	30 - 44	34.0	37.0	650.3		36.6	17.0		37.5	120.5	
	45 - 59	25.3	23.5	(<.001)		24.8	(<.001)		30.0	(<.001)	
	60 - 74	15.6	13.8			14.1			17.7		

NS = Non Significant

Table 6.3.2
Representativeness: psychotropic medication prescription in the GP practices

Prescription	GP population N = 80,315 %	GP sample N = 4,517 %	χ^2	(p)	Survey sample N = 1,813 %	χ^2	(p)
Antidepressants	3.4	3.0	2.0	NS	3.4	0.0	NS
Benzodiazepines	9.5	8.7	3.0	NS (.08)	10.3	1.6	NS
Antipsychotics	0.8	0.6	2.8	NS (.10)	0.4	3.3	NS (.07)
Analgesics	9.6	9.6	0.0	NS	10.3	0.9	NS
Other psychotropic drugs	2.7	2.7	0.0	NS	2.7	1.4	NS

NS = Non Significant

6.3.2 Psychiatric prevalence

6.3.2.1 Prevalence: full diagnostic range

In the logistic regression model, α was estimated at -2.52 and β at +0.16. After correction for gender, age and urbanisation, the 1-month overall prevalence rate of psychiatric morbidity in the Nijmegen Health Area turned out to be 17.6% (± 2.6%). Variables such as gender, age, and urbanisation did not improve the model.

The prevalence rates for the respective diagnostic categories and groups are reported in Table 6.3.3. The four categories with the highest prevalence rates were sleep (4.8%), anxiety (4.6%), substance-related (4.0%), and mood disorders (3.8%). Substance abuse (3.2%), depression (2.7%), and phobic or panic disorders (2.6%) were the diagnostic groups most frequently found.

6.3.2.2 Prevalence: influence of the three newly determined categories

When the three newly determined diagnostic categories in the general population (sleep, dissociative, and somatoform disorders) were left out of the logistic regression model, α was estimated at -2.99 and β at +0.15. In comparison with the full diagnostic range, the prevalence rate of psychiatric caseness decreased by a third, from 17.6% to 11.9% (±2.5%).

Table 6.3.3
Prevalence rates* of psychiatric disorders with reference to caseness, categories and groups

	Prevalence		Prevalence
Caseness	17.6 (± 2.6) †		
Psycho-organic disorder	0.3	Eating disorder	0.4
Dementia	0.3	Anorexia Nervosa	0.1
Other organic brain disorder	0.0	Other eating disorders	0.3
Substance-related disorder	4.0	Somatoform disorder	2.8
Dependence	0.9	Somatisation disorder	0.1
Abuse	3.2	Pain disorder	1.2
Psychotic disorder	0.5	Other somatoform disorder	1.5
Schizophrenia	0.4	Dissociation disorder	1.4
Delusional disorder, non-aff. psychosis	0.1	Sleep disorder	4.8
Mood disorder	3.8	Dyssomnia	2.2
Depression	2.7	Parasomnia	2.1
Mania, hypomania, mixed disorder	0.3	Other sleep disorder	0.5
Dysthymic disorder	1.9	Other psychiatric disorder	0.7
Anxiety disorder	4.6		
Generalised anxiety disorder	0.3		
Phobic or panic disorder	2.6		
Obsessive-compulsive disorder	1.3		
Other anxiety disorders	0.6		

* after correction for gender, age, and urbanisation; due to comorbidity, cumulative percentages of diagnostic groups within a given category may exceed percentages for diagnostic categories

† 95% confidence interval

6.3.3 Distribution of psychiatric disorders

6.3.3.1 Distribution: full diagnostic range

The sociodemographic distribution of psychiatric caseness is presented in Table 6.3.4. The prevalence rate of caseness was significantly higher for females (19.1%) than for males (16.2%, Table 6.3.5). This also applied to subjects living in the city of Nijmegen (20.1%) in comparison with those resident in the surrounding communities (16.3%). Table 6.3.4 shows that the age group 45-54 yrs (20.1%), divorced subjects (22.7%), unemployed subjects (29.5%) and chronically ill subjects (29.5%), and those in a lower education bracket (20.0%) had a higher caseness rate than the subjects in other sociodemographic subgroups. Apart from the age effect, this distribution agreed with the findings of Dohrenwend and Dohrenwend (1974).

The distribution of diagnostic categories across gender, age, and urbanisation subgroups is presented in Table 6.3.5. The number of age groups was reduced to obtain sufficient power for the analysis of the more detailed diagnostic categories and groups.

The rates for females were equal to or higher than those for males for all diagnostic categories and groups, except for substance-related disorders, for which the rate was significantly higher for males, mainly because of a difference in substance abuse rates. Mood disorders, in particular depression and dysthymia, were two to three times more common amongst females, while the prevalence rate of sleep disorders, a DSM category rarely assessed in the general population, was three to four times higher for females than for males.

The prevalence rate of substance-related disorders decreased with increasing age, with a significantly lower rate of substance abuse for the oldest subjects. A tendency toward the highest rates for mood disorders was found in the age group 45-59 yrs, especially due to a significant contribution of depressive disorders, which was significantly lower in younger persons. The rates of sleep disorders were also lower for younger people.

The prevalence rate of caseness for subjects living in the city of Nijmegen was higher than for those resident in the surrounding communities. Nearly all diagnoses, especially substance-related disorders and mood disorders, were more frequent in the city of Nijmegen than in the surrounding communities, except for somatoform disorders[vi], for which the prevalence rate was higher in the surrounding communities.

Table 6.3.4
Sociodemographic distribution of psychiatric caseness

	Total		Males				Females			
			Nijmegen		SC *		Nijmegen		SC *	
	N	%	N	%	N	%	N	%	N	%
Caseness	1,799	17.6	214	18.8	589	14.8	311	21.3	685	17.8
Age group (yrs)										
18-24	120	16.0	16	21.9	28	12.4	29	19.6	47	13.9
25-34	359	18.4	44	18.5	93	13.2	86	20.9	136	20.4
35-44	461	16.8	64	17.9	151	15.2	80	19.8	166	16.5
45-54	391	20.1 [a]	41	20.7	144	18.1	53	25.3	153	20.0
55-64	272	16.7	26	19.8	102	15.2	28	26.9	116	14.9
≥ 65	194	14.7 [b]	23	14.9	70	12.7	35	17.6	66	15.2
Marital status										
Single	141	19.8	41	22.2	33	20.8	48	19.1	19	14.9
Living with parents	68	15.4	5	24.5	32	13.2	3	25.0	28	15.4
Living with partner	214	18.2	35	17.9	39	17.0	84	20.1	56	16.5
Married	1,209	17.2	117	18.1	454	14.4	124	22.6	514	18.2
Divorced	88	22.7	11	15.2	24	20.6	31	27.3	22	22.2
Widowed	77	16.9	5	14.2	6	19.3	21	16.3	45	17.2
Employment status**										
Full-time employed	706	15.6 [b d]	130	16.2	384	14.4	92	16.2	100	18.7
Part-time employed	378	18.0 [b d]	23	15.1	27	14.3	90	21.6	238	17.4
Retired	189	14.8 [b d]	28	15.6	100	14.7	20	20.8	41	11.3
Chronically ill/unable to work	74	29.5 [a e]	6	58.8	42	21.5	9	36.3	17	35.4
Unemployed	40	29.5 [c]	10	35.0	12	20.3	9	39.5	9	25.6
School or college student	60	17.8 [b]	12	24.6	14	15.4	13	19.3	21	14.6
Householder	285	19.9 [f]	--	--	1	7.5	57	26.9	227	18.2
Others without an occupation	59	17.4 [b]	5	25.7	7	16.4	19	16.0	28	17.1
Educational level										
University	151	17.4	46	18.7	35	16.0	62	17.1	8	18.4
Higher vocational education	314	15.3	36	16.8	97	14.0	82	16.8	99	14.7
Upper-stream sec. education	162	17.9	36	19.5	40	10.9	35	21.7	51	19.7
Lower-stream sec. & post-sec. lower vocational education	338	16.0	26	17.2	130	14.6	16	24.4	166	16.1
Lower-stream sec. education	226	18.5	18	17.9	55	16.7	32	20.5	121	18.8
Continued primary education	423	19.7	38	21.9	156	15.7	53	28.7	176	20.0
No education or only primary education	183	20.0	14	17.9	75	17.5	31	28.9	63	18.9

* SC: surrounding communities

** N = 6 missing

Age group: Kruskal-Wallis .0001 < p < .001; Scheffé (.05) [a] – [b]

Marital status: Kruskal-Wallis .001< p < .01; Scheffé, non-significant

Employment status: Kruskal-Wallis p < .0001; Scheffé (.05) [a] – [b], [c] – [d], and for females only [e] - [f]

Level of education: Kruskal-Wallis .001< p< .01; Scheffé, non-significant

Table 6.3.5
Distribution[a] of psychiatric caseness, diagnostic categories and groups across gender, age and urbanisation (legend next page)

	Gender			Age					Urbanisation		
	Male N=803	Female N = 996	p	18-29 N = 269	30-44 N=671	45-59 N = 541	60-74 N = 318	p	City N=525	SC* N=1,274	p
Caseness	16.2	19.1	***	17.1	17.2	20.1[b]	15.5[c]	***	20.1	16.3	***
Substance-related disorder	6.8	1.8	***	5.5[d]	4.7	4.2	0.4[e]	*	6.9	3.0	**
Dependence	1.0	0.9		1.4	1.4	0.5	0.0		1.8	0.5	**
Abuse	5.9	0.9	***	4.1[d]	3.4	3.7	0.4[e]	*	5.1	2.5	
Psychotic disorder	0.5	0.5		1.7	0.4	0.2	0.0		1.4	0.2	
Schizophrenia	0.3	0.5		1.7	0.2	0.2	0.0		1.4	0.1	
Mood disorder	2.2	5.1	*	2.3	3.7	4.8	3.5		6.6	2.7	**
Depression	1.3	3.9	**	1.4[e]	2.3	4.1[d]	2.6	*	4.7	2.0	*
Mania, hypomania, mixed disorder	0.3	0.3		0.0	0.4	0.5	0.0		0.0	0.4	
Dysthymic disorder	1.0	2.7	*	1.4	1.8	2.6	1.3		3.7	1.2	*
Anxiety disorder	3.0	6.0		4.9	4.8	4.8	3.0		5.0	4.3	
Generalised anxiety disorder	0.2	0.4		0.5	0.2	0.2	0.4		0.5	0.2	
Phobic or panic disorder	1.8	3.2		1.8	3.2	2.6	1.3		1.8	2.8	
Obsessive-compulsive disorder	0.8	1.8		2.2	1.1	1.4	0.9		1.3	1.3	
Other anxiety disorder	0.3	0.9		0.5	0.5	1.0	0.4		1.3	0.4	
Eating disorder	0.0	0.6		0.0	0.7	0.2	0.0		0.5	0.3	
Anorexia Nervosa	0.0	0.1		0.0	0.2	0.0	0.0		0.3	0.0	
Other eating disorder	0.0	0.5		0.0	0.5	0.2	0.0		0.3	0.3	
Somatoform disorder	2.8	2.8		2.7	2.9	2.4	3.1		2.4	2.9	
Somatisation disorder	0.2	0.1		0.5	0.0	0.2	0.0		0.5	0.0	*
Pain disorder	1.0	1.4		0.9	1.1	0.7	2.7		1.6	1.0	
Other somatoform disorder	1.8	1.3		1.4	1.9	1.7	0.4		0.5	1.9	*
Dissociation disorder	1.0	1.8		0.5	1.7	2.1	0.0		3.5	0.7	
Sleep disorder	2.1	7.1	***	1.8[e]	4.3[e]	6.7[d]	5.3	**	4.5	4.8	
Dyssomnia	0.6	3.5	**	0.0	2.5	2.6	2.7		1.6	2.4	
Parasomnia	0.8	3.2	**	0.5[e]	1.8	3.3[d]	2.2	*	2.6	1.9	
Other sleep disorder	0.6	0.4		1.4	0.0	0.7	0.4		0.3	0.6	
Other psychiatric disorder	1.0	0.4		0.5	0.2	1.4	0.9		1.9	0.3	

Legend to table 6.3.5

* SC: surrounding communities

[a] after correction for gender, age, and urbanisation

p: Caseness: ANOVA: * = .01 < p ≤ .05; ** = .001 < p ≤ .01; *** = p ≤ .001, Scheffé (.05) [b] - [c]

Categories and groups: χ^2: * = .01 < p ≤ .05; ** = .001 < p ≤ .01; *** = p ≤ .001; for age: Scheffé (.05) [d] - [e]

6.3.3.2 Distribution: influence of the three newly determined categories

When the three newly determined categories were left out, the sociodemographic distribution of caseness was characterised by about the same pattern as the distribution of caseness for the full diagnostic range (Table 6.3.6). At first sight, the geographical distribution of two of the three categories appears not to correspond to the findings of Dohrenwend and Dohrenwend, as the prevalence rates for somatoform disorders and sleep disorders were higher in the surrounding communities than in Nijmegen (Table 6.3.5). But when the hierarchical rules of DSM-IV were not applied, the regional distribution of the prevalence rates corresponded to its expected pattern of a higher rate in the city than in the surrounding communities (somatoform disorders 3.9% and 3.4%, and sleep disorders 6.8% and 5.8%).

Table 6.3.6
Sociodemographic distribution of psychiatric caseness without the three newly determined categories

	Total		Males				Females			
			Nijmegen		SC*		Nijmegen		SC*	
	N	%	N	%	N	%	N	%	N	%
Caseness	1,799	11.9	214	12.9	589	9.8	311	14.8	685	12.0
Age group (yrs)										
18-24	120	10.5	16	15.1	28	7.9	29	13.1	47	8.9
25-34	359	12.4	44	12.2	93	8.5	86	14.4	136	13.9
35-44	461	11.3	64	12.3	151	10.1	80	13.6	166	11.0
45-54	391	14.4 [a]	41	14.4	144	12.4	53	17.8	153	15.2
55-64	272	11.3	26	14.3	102	10.2	28	19.4	116	9.7
≥ 65	194	9.8 [b]	23	10.0	70	8.4	35	12.0	66	10.0
Marital status										
Single	141	13.2	41	15.0	33	14.0	48	12.6	19	9.5
Living with parents	68	10.1	5	17.4	32	8.3	3	17.2	28	10.0
Living with partner	214	12.3	35	11.9	39	11.6	84	13.8	56	10.1
Married	1,209	11.7	117	12.7	454	9.5	124	15.9	514	12.3
Divorced	88	15.6	11	9.8	24	14.6	31	19.3	22	15.3
Widowed	77	11.2	5	8.9	6	12.5	21	11.1	45	11.3
Employment status**										
Full-time	706	10.2 [b]	130	10.6	384	9.5	92	10.6	100	12.4
Part-time	378	12.1 [b]	23	9.9	27	9.5	90	14.8	238	11.6
Retired	189	10.0 [b]	28	10.6	100	10.0	20	14.5	41	7.2
Chronically ill/unable to work	74	21.6 [a c]	6	47.3	42	14.9	9	29.1	17	25.2
Unemployed	40	21.7 [a]	10	27.2	12	14.0	9	28.7	9	18.9
School or college student	60	11.8 [b]	12	17.3	14	9.8	13	12.8	21	9.4
Householder	285	13.7 [d]	--	--	1	4.8	57	19.3	227	12.3
Others without an occupation	59	12.0 [b]	5	19.2	7	10.6	19	11.2	28	11.5
Educational level										
University	151	11.5	46	12.7	35	10.6	62	11.2	8	11.8
Higher vocational education	314	10.0	36	11.0	97	9.2	82	11.1	99	9.6
Higher secondary education	162	12.0	36	13.3	40	7.0	35	14.5	51	13.3
Lower-stream sec. & post-sec. lower vocational education	338	10.7	26	11.2	130	9.7	16	18.1	166	10.7
Lower-stream sec. education	226	12.6	18	12.5	55	10.9	32	14.5	121	12.9
Continued primary education	423	13.6	38	15.9	156	10.6	53	20.8	176	13.7
No education or only primary education	183	13.6	14	12.9	75	11.8	31	20.8	63	12.4

* SC: surrounding communities

** N = 6 missing

Age group: Kruskal-Wallis: .0001 < p< .001; Scheffé (.05) [a] – [b]

Marital status: Kruskal-Wallis: .001< p< .01; Scheffé, non-significant

Employment status: Kruskal-Wallis: p< .0001; Scheffé (.05) [a] – [b] and for females only [c] - [d]

Educational level: Kruskal-Wallis: .001< p< .01; Scheffé, non-significant

6.3.4 Psychiatric comorbidity

6.3.4.1 Comorbidity: full diagnostic range

Although DSM-IV includes a diagnostic hierarchy, one person can have more than one diagnosis. Of the total number of cases, 26.1% had more than one diagnosis (17.1% two, 4.1% three, and 4.9% four or more diagnoses - see Table 6.3.7). The comorbidity rates for females were higher than for males, which was significant for subjects with two diagnoses. For all numbers of diagnoses per subject, prevalence rates of comorbidity differed significantly for age and urbanisation. Comorbidity rates were higher for subjects aged 45 to 59 years and subjects living in the city of Nijmegen.

Table 6.3.7
Distribution of diagnostic comorbidity across gender, age and urbanisation

	Total	Gender			Age					Urbanisation		
		male N=803	female N=996	p	18-29 N=269	30-44 N=671	45-59 N=541	60-74 N=318	p	city N=525	SC N=1,274	p
Caseness	17.6	16.2	19.1	***	17.1	17.2	20.1 [a]	15.5 [b]	***	20.1	16.3	***
1 diagnosis	13.0	12.3	13.8	**	12.7	12.8	14.4 [c]	11.9 [d]	***	14.4	12.3	***
2 diagnoses	3.0	2.7	3.3	**	2.7	2.9	3.6 [c]	2.7 [d]	**	3.6	2.7	***
3 diagnoses	0.7	0.6	0.8		0.5	0.7	1.0	0.6	*	1.1	0.5	***
4 diagnoses	0.9	0.7	1.0		0.6	0.8	1.3	0.8	*	1.3	0.6	***

Caseness: ANOVA: * = .01 < p ≤ .05; ** = .001 < p ≤ .01; *** = p ≤ .001, Scheffé (.05) [a] - [b]

Number of diagnoses: χ^2: * = .01 < p ≤ .05; ** = .001 < p ≤ .01; *** = p ≤ .001; for age: Scheffé (.05) [c] - [d]

6.3.4.2 Comorbidity: influence of the three newly determined categories

The cumulative prevalence rate of the three categories was 9% (Table 6.3.3), whereas their actual contribution to caseness was only 5.7% (17.6% and 11.9% with and without the three categories). This means that the comorbidity rate for the three newly determined categories themselves was 36.7% (Table 6.3.8), as against 26.1% for comorbidity for the full diagnostic range. As a result, of the total number of subjects who constituted a case when the three categories were left out, only 22.3% were comorbid.

Table 6.3.8
Distribution of diagnostic comorbidity across gender, age and urbanisation, without the three newly determined categories

	Total	Gender			Age					Urbanisation		
		male N=803	female N=996	p	18-29 N=269	30-44 N=671	45-59 N=541	60-74 N=318	p	city N=525	SC N=1,274	p
Caseness	11.9	10.8	13.0	***	11.3	11.6	13.8 [a]	10.4 [b]	***	13.9	10.9	***
1 diagnosis	9.3	8.7	9.8	**	9.0	9.1	10.3 [c]	8.4 [d]	***	10.3	8.7	***
2 diagnoses	1.6	1.4	1.8	**	1.4 [d]	1.6	2.1 [c]	1.4 [d]	**	2.1	1.4	***
3 diagnoses	0.7	0.5	0.8		0.4	0.6	1.0	0.6	*	1.0	0.5	***
4 diagnoses	0.4	0.3	0.4		0.2	0.3	0.6	0.3	*	0.6	0.2	***

Caseness: ANOVA: * = .01 < p ≤ .05; ** = .001 < p ≤ .01; *** = p ≤ .001, Scheffé (.05) [a] - [b]

Number of diagnoses per subject: χ^2: * = .01 < p ≤ .05; ** = .001 < p ≤ .01; *** = p ≤ .001; for age: Scheffé (.05) [c] - [d]

6.3.5 The NHA-2 and the NEMESIS

The first wave of the Netherlands Mental Health Survey and Incidence Study (NEMESIS) was carried out in 1996 with the aid of the CIDI-1.1 (Bijl et al. 1998). A total of five main diagnostic categories were included in the psychiatric interview in the general population, aged 18 to 64 years.

When the survey sample (with respect to age) and diagnostic range in the NHA-2 were reduced to the extent adopted in the NEMESIS, the prevalence rate for the NHA-2 was 11.7% (±2.2%), which was substantially lower than the 16.5% (±0.4%) prevalence rate in the NEMESIS study (see Table 6.3.9). Yet the distribution of the prevalence rates by gender is comparable for the two studies, as in both the prevalence rates were higher for females for all diagnostic categories and groups, except for the substance use-related disorders.

The most striking difference between the two studies from a diagnostic perspective (probably due to a difference in the clinical experience of the interviewers) was in the anxiety category, especially in the phobic groups. The prevalence rate for this category was twice as high in the NEMESIS study as in the NHA-2 (9.7% vs. 4.9%), the biggest difference occurring in the specific phobia group (5.5% vs. 0.8%). As the only exception, the prevalence rate of obsessive-compulsive disorder (OCD) in the NHA-2 was higher than in the NEMESIS study (1.4% vs. 0.3%).

Another marked difference was found with regard to substance-related disorders, for which the prevalence rate in the NHA-2 was lower than in the NEMESIS (4.4% vs. 5.8%), and for males the rate of substance abuse was higher and that of dependence lower than in the NEMESIS.

The rate of psychotic disorders found in the NEMESIS study was lower than in the NHA-2 study (0.2% vs. 0.5%).

Table 6.3.9
Prevalence rates [a] in the NHA-2 and the NEMESIS study

	NHA-2			NEMESIS		
	Male	Female	Total	Male	Female	Total
	N = 710	N = 895	N = 1,605	N = 3,777	N = 3,299	N = 7,076
Caseness	10.7	12.7	11.7	15.9	17.0	16.5
Substance use-related disorder	7.4	1.9	4.4	9.2	2.4	5.8
Dependence	1.0	1.0	1.0	5.1	1.4	3.3
Abuse	6.4	1.0	3.4	4.2	1.0	2.6
Psychotic disorder	0.5	0.5	0.5	0.1	0.2	0.2
Schizophrenia	0.3	0.5	0.4	---	---	---
Delusional disorder, non-aff. psychosis	0.2	0.0	0.1	---	---	---
Mood disorder	2.4	5.2	4.0	2.8	5.0	3.9
Depression	1.4	4.1	2.9	1.9	3.4	2.7
Mania, hypomania, mixed disorder	0.3	0.3	0.3	0.4	0.8	0.6
Dysthymic disorder	1.0	2.8	2.0	1.0	2.1	1.6
Anxiety disorder	3.1	6.4	4.9	6.5	12.9	9.7
Generalised anxiety disorder	0.2	0.4	0.3	0.6	1.0	0.8
Panic disorder	0.1	0.5	0.3	0.8	2.2	1.5
Agoraphobia (without panic disorder)	0.0	0.4	0.2	0.6	1.4	1.0
Specific phobia	0.2	1.3	0.8	3.1	8.0	5.5
Social phobia	1.5	1.1	1.3	2.8	4.7	3.7
Obsessive-compulsive disorder	0.9	1.8	1.4	0.3	0.2	0.3
Eating disorder	0.0	0.7	0.4	0.1	0.4	0.3
Anorexia Nervosa	0.0	0.1	0.1	0.0	0.0	0.0
Other eating disorders	0.0	0.6	0.3	0.1	0.4	0.3

[a] after correction for gender, age, and urbanisation; due to comorbidity, cumulative percentages of diagnostic groups within a given category may exceed percentages for diagnostic categories.

6.4 Discussion

6.4.1 Study limitations

The fact that the respondents were recruited as a random sample from the patient registers of GPs had two consequences. The first consequence was an advantage: we were able to check for selectivity by means of the prescription rate for psychotropic medication. The second consequence was a disadvantage: the medical ethics committee deemed that the epidemiological part of the study was no longer a field survey in the general population, but a patient-oriented study. As a result, potential respondents seemed to be deterred by the exhaustive informed consent letter judged to be necessary by the medical ethical committee. Consequently, the response rate did not meet our expectations. Nevertheless, with respect to the survey, the sample turned out to be representative of the general population as far as psychiatric morbidity was concerned. That was shown by the fact that there were neither significant, nor relevant differences in the prescription rates for psychotropic drugs, which was adopted as an indication of mental illness. The non-response between phases T1 and T2 (19%) corresponded to that of other two-phase studies in the general population (McConnell et al. 2002; Roca et al. 1999). The choice of an expensive and demanding clinical interview instrument naturally imposed limitations on the number of interviewees in comparison with the number in the NEMESIS and the National Comorbidity Study (NCS) study (Bijl et al. 1998; Kessler et al. 1994; Narrow et al. 2002).

6.4.2 Psychiatric prevalence

In the NHA-2 study, we found the **overall prevalence rate** for psychiatric disorders in the general population to be 17.6%. Prevalence rates for psychiatric morbidity vary considerably (WHO 2000). In their overview of a number of larger studies in the general population, Goldberg and Huxley (1992) cite a mean 1-month prevalence rate of 16.4%, for which the range is large (7.3% (Hodiamont et al. 1987) to 26.2% (Cheng 1988)). Their data are based on the PSE and the Diagnostic Interview Schedule (the DIS, the precursor of the CIDI), which have a limited diagnostic range in comparison with later interviews. The prevalence rate in the NHA-2 for the incomplete range was 11.9%, which was 5.8% lower than for the complete range. This implies that the estimated 1-month complete prevalence rate cited by Goldberg and Huxley would increase to 22.2% if the integral diagnostic range were adopted.

The NHA-2 study is the first epidemiological survey in the general population in the Netherlands to have used a semi-structured, clinically oriented interview (SCAN-2.1) covering the complete range of DSM-IV disorders (with the

exception of disorders usually first diagnosed in infancy, childhood, or adolescence). From this perspective, the overall prevalence rate of 17.6%, although significantly higher than the rate determined from the 1983 NHA-1 data (see chapter 5), can be considered as below average, given the contribution of sleep, somatoform, and dissociative disorders, which usually are not assessed. (Angst et al. 2005; Bijl et al. 1998; Kessler et al. 1994; Kessler et al. 2005b). One important reason for this finding is the clinical nature of the SCAN-2.1 (Brugha et al. 1999a; Brugha et al. 2001; Cooper & Singh 2000; Goldberg & Huxley 1992; Narrow et al. 2002).

6.4.3 Distribution of psychiatric disorders

The **demographic distribution o**f the diagnoses in the NHA-2 study constitutes a recognisable picture (Dohrenwend & Dohrenwend 1974), with higher rates for females than for males for most disorders, with the exception of substance-related disorders. In agreement with other studies, higher rates were found in the city than in the surrounding communities (Glover et al. 1998; Hodiamont et al. 1992; Jenkins et al. 1997; Peen et al. 2002). On the other hand, there is no consensus in the literature on the relationship between age and psychiatric morbidity. Hodiamont et al. (1987) reported a general increase in prevalence rates with increasing age, whereas other investigators state that the rates tend to decrease with age. Goldberg and Huxley (Goldberg & Huxley 1992) suggested that “generally speaking, rates for common mental disorders peak in the middle years”. In our study, we found a significant peak in the prevalence rates for persons aged 45 to 54 years (Table 6.3.4 and Table 6.3.6) and a lower rate for the highest age group, despite the fact that psycho-organic disorders were part of our diagnostic spectrum. With regard to educational level and marital status, our findings are consistent with those reported in the literature in that the prevalence rate for caseness for subjects with little education, and unemployed, chronically ill and divorced subjects was higher.

6.4.4 Psychiatric comorbidity

In the NHA-2 study, which was based on the SCAN-2.1, the comorbidity rate of the subjects with a diagnosis was substantial (26%), but lower than in most CIDI studies. The overall 1-month comorbidity rate for the NEMESIS was slightly higher, namely just over 30% (Bijl et al. 1998; Ravelli et al. 1998). Roca (1999) reports the 1-month comorbidity rate of the Formentera study to be 36.2%. One-month comorbidity data from the NCS study are not available. With the prolongation of the period of reference, comorbidity rates in subjects with at least one diagnosis increase to 56% (NCS) and even 60% (ECA) when lifetime is used as a reference (Bijl et al. 1998; Kessler et al. 1994; Ravelli et al. 1998;

Robins et al. 1991). Kessler (2005b) reports the 12-month comorbidity rate of the NCS-Replication study to be 45%.

6.4.5 Influence of the three newly determined categories

The three psychiatric diagnostic categories newly determined in the general population contributed substantially to the psychiatric prevalence rate, which increased by almost 50% when they were included. Bebbington et al. (1997) reported a contribution of sleep disorders to the prevalence rate amounting to almost one third.

The three categories had the same sociodemographic distribution as most other diagnostic categories, apart from the somatoform disorders, for which there was no difference in distribution between the sexes, and a higher prevalence rate for the elderly and the rural dwellers. One important explanation of the deviation from the regular pattern of the sociodemographic distribution of these disorders (no gender difference and a higher prevalence rate for rural dwellers) is that pain disorders (somatoform disorder) and dyssomnia (sleep disorder) in particular share key symptoms with major mood and anxiety disorders. Because mood and anxiety disorders prevail hierarchically over dyssomnia and pain disorders, the latter two disappear. As a consequence, higher prevalence rates for mood and anxiety disorders in females and city people have a reducing effect on the prevalence rates of these two newly determined categories. When the hierarchical rules of DSM-IV are omitted, these differences disappear and the expected pattern of higher rates for females and city dwellers is observed.

The additional three newly determined categories also contributed to the comorbidity rate (36.7%) to a greater extent than the commonly determined part of the diagnostic range (22.3%). Furthermore, as these categories are partially hidden as a result of the DSM hierarchical rules, comorbidity would be much higher still if these rules were not taken into account. This illustrates their importance to this health issue, since, on the one hand, comorbidity points to the complexity of the underlying problems, and on the other, most interventions (medication, psychotherapy, etc.) are not comorbidity ‘proof’, in that they are focussed on the treatment of well-defined monomorbid disorders.

6.4.6 The NEMESIS

The **NHA-2 study and the NEMESIS** are two epidemiological studies on psychiatric morbidity in the general population, the two of them having been carried out in the Netherlands at about the same time. Because of the co-occurrence of the two studies, it is meaningful to compare their results with one another. The comparability of the NEMESIS and the NHA-2 results is hindered by a number of differences between the two designs. In particular, the

NEMESIS data were collected from a nation-wide sample of the general population aged 18 through 64 years, in one phase, and throughout 1996, with the CIDI, a fully structured interview based on the DSM-III-R with a limited diagnostic range, which was administered by trained lay interviewers. In contrast, the NHA-2 addressed a regional sample of the general population aged 18 through 75 years, from September 1997 to February 1998, on the basis of a two-stage, two-phase design, with the aid of the SCAN-2.1, a semi-structured, clinical interview based on the integral diagnostic range of DSM-IV and administered by trained clinicians. For two of the differences (age and diagnostic range), the NHA-2 data could be adjusted before comparison. All things considered, the differences could be attributed to differences in the **diagnostic instrument (CIDI vs. SCAN)** that was used, the **classification system (DSM-III-R vs. DSM-IV)** that was adopted, and the **geographical area (national vs. regional)** in which the sampling was done.

The two studies generated markedly different results. The higher overall prevalence rate found in the NEMESIS (16.5% vs. 11.7% in the NHA-2) can to a great extent be accounted for by the anxiety (i.e., specific phobia) rates. Specific phobia is one of the diagnoses that are based on the subject's account of his or her experiences in everyday life situations.

Unlike the **SCAN** ratings, the **CIDI** ratings are based purely on the subjective respondent's answers ('yes' or 'no'), which tend to be influenced by *subjective* feelings of perceived morbidity (*illness),* no account being taken of a judgement about the clinical relevance of each experience, sign or symptom separately. Given the preceding, it is conceivable that comparable items are scored more readily in the CIDI than in the SCAN, resulting in a higher prevalence rate (Brugha et al. 1999a; Brugha et al. 2001; Cooper & Singh 2000).

The fact that the rate of substance-related disorders was lower in the NHA-2 than in the NEMESIS has more than one explanation. One of these is the difference in the criteria used in the **DSM-III-R** (NEMESIS) and the **DSM-IV** (NHA-2): the DSM-III-R requires at least three out of *nine* symptoms, whereas the DSM-IV requires three out of *seven*[vii], the latter being harder to satisfy statistically. Furthermore, there is a difference between the two manuals with regard to the period of reference: this is "a longer period than the person intended" in DSM-III-R, and a fixed 12-month period in the DSM-IV, the former literally introducing *subjectivity* as the point of reference. As a result of these differences, the higher *dependence* rate in the NEMESIS reduces the *abuse* rate because of the hierarchical relationship in the DSM between dependence and abuse (the existence of *dependence* is an exclusion criterion for *abuse* in both manuals). Another explanation of the lower dependence rate in the NHA-2 might be that the NEMESIS, being a **national study**, included metropolitan areas, which possibly contributed to a higher rate of these disorders than in the NHA-2 study.

6.4.7 Conclusions

All in all, we deem the SCAN-2.1 to be a clinical instrument feasible for use in general population surveys, its advantage being that clinical standards can be maintained. Comparison of the results of the NHA-2 with those of the NEMESIS supports our conclusion that clinical interviews are suitable for testing the psychiatric (clinical) relevance of common signs and symptoms, and consequently lead to lower and more plausible prevalence rates (Brugha et al. 2001).

The use of the SCAN-2.1 and the integral range of the DSM-IV resulted in a prevalence rate of 17.6%, to which the diagnostic categories newly determined in the general population, such as sleep, dissociative, and somatoform disorders, contributed a substantial part, namely 5.7%. The sociodemographic distribution of psychiatric disorders turned out to be as Dohrenwend and Dohrenwend described it years ago.

Despite the use of the complete diagnostic range, comorbidity rates in the NHA-2 can be classified as moderate, notwithstanding the boosting effect of the three newly determined diagnostic categories.

Fully structured interviews like the ones used in the NEMESIS and the NCS yield higher caseness rates than clinical interviews like the SCAN-2.1. Although the literature on general population studies performed with the SCAN is still limited, it may be concluded that the NHA-2 has the most moderate levels of psychiatric caseness to date.

With regard to costs, the CIDI makes general population studies more feasible than the SCAN, but currently gives rise to improbably high prevalence rates for specific diagnostic categories. Because of the large discrepancies in prevalence rates between the SCAN and the CIDI for diagnostic caseness and also diagnostic categories, the use of the former is preferable as far as methodological validity is concerned. But it has its price and the challenge before us is to *improve* the CIDI in this respect. More direct comparative studies between the SCAN and the CIDI (as presented in chapter 9) will be needed to achieve an understanding of how best to incorporate clinical relevance into the CIDI.

References

Andrews G, Peters L, Guzman AM and Bird K. (1995) A comparison of two structured diagnostic interviews: CIDI and SCAN. *Australian and New Zealand Journal of Psychiatry,* 29, 124-132.

Angst J, Gamma A, Neuenschwander M, Ajdacic-Gross V, Eich D, Rossler W and Merikangas KR. (2005) Prevalence of mental disorders in the Zurich Cohort Study: a twenty year prospective study. *Epidemiologia e Psichiatria Sociale,* 14, 68-76.

Batstra L, Bos EH and Neeleman J. (2002) Quantifying psychiatric comorbidity: lessons from chronic disease epidemiology. *Social Psychiatry and Psychiatric Epidemiology,* 37, 105-111.

Bebbington PE, Marsden L and Brewin CR. (1997) The need for psychiatric treatment in the general population: the Camberwell Needs for Care Survey. *Psychological Medicine,* 27, 821-834.

Bijl RV, Ravelli A and van Zessen G. (1998) Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Social Psychiatry and Psychiatric Epidemiology,* 33, 587-595.

Boersma WGW, de Jong FAJM and Mulder PH. (1993) The Netherlands. In *Health Care and General Practice across Europe*. pp. 77-80. Nivel (Netherlands Institute of Primary Health Care), Utrecht.

Brugha T, Bebbington P, Tennant C and Hurry J. (1985) The List of Threatening Experiences: a subset of 12 life event categories with considerable long-term contextual threat. *Psychological Medicine,* 15, 189-194.

Brugha TS, Bebbington PE and Jenkins R. (1999a) A difference that matters: comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population [editorial]. *Psychological Medicine,* 29, 1013-1020.

Brugha TS, Jenkins R, Taub N, Meltzer H and Bebbington PE. (2001) A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychological Medicine,* 31, 1001-1013.

Brugha TS, Nienhuis F, Bagchi D, Smith J and Meltzer H. (1999b) The survey form of SCAN: the feasibility of using experienced lay survey interviewers to administer a semi-structured systematic clinical assessment of psychotic and non-psychotic disorders. *Psychological Medicine,* 29, 703-711.

CBS. (1997) *Census data 1997*. Voorburg/Heerlen.

Cheng AT, Tien AY, Chang CJ, Brugha TS, Cooper JE, Lee CS, Compton W, Liu CY, Yu WY and Chen HM. (2001) Cross-cultural implementation of a Chinese version of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) in Taiwan. *British Journal of Psychiatry,* 178, 567-572.

Cheng TA. (1988) A community study of minor psychiatric morbidity in Taiwan. *Psychological Medicine,* 18, 953-968.

Cooper B and Singh B. (2000) Population research and mental health policy. Bridging the gap [editorial]. *British Journal of Psychiatry,* 176, 407-411.

de Jong-Gierveld J and van Tilburg T. (1993) *Manual of the Loneliness scale*. Department of Social Research Methodology, VU Amsterdam, Amsterdam.

Dohrenwend BP and Dohrenwend BS. (1974) Social and cultural influences on psychopathology. *Annual Review of Psychology,* 25, 417-452.

Duine TJ, Hodiamont PPG, Peer PGM and Boerma LH. (1985) Reliability aspects of the Present State Examination in four different populations [in Dutch]. *Tijdschrift voor Psychiatrie,* 27, 326-340.

Duncan-Jones P. (1979) Validity and uses of the GHQ [letter]. *British Journal of Psychiatry,* 135, 382.

Dunn G, Pickles A, Tansella M and Vazquez-Barquero JL. (1999) Two-phase epidemiological surveys in psychiatric research [editorial]. *British Journal of Psychiatry,* 174, 95-100.

Eaton WW, Neufeld K, Chen LS and Cai G. (2000) A comparison of self-report and clinical diagnostic interviews for depression: diagnostic interview schedule and schedules for clinical assessment in neuropsychiatry in the Baltimore epidemiologic catchment area follow-up. *Archives of General Psychiatry,* 57, 217-222.

Efron B and Tibshirani RJ. (1993) *Introduction to the bootstrap*. (Vols. 57) Chapman and Hall, New York.

Feinstein AR. (1970) The pre-therapeutic classification of comorbidity in chronic disease. *Journal of Chronic Diseases,* 23, 455-468.

Furer JW, König-Zahn C and Tax B. (1995) General Health Questionnaire (GHQ) [in Dutch]. In JW Furer, C König-Zahn en B Tax (Eds.), *The measurement of health*. pp. 104-118. van Gorcum, Assen.

Giel R and Nienhuis FJ. (1997) *SCAN-2.1: Schedules for clinical assessment in neuropsychiatry [in Dutch]*. WHO, Geneva / Groningen.

Glover GR, Robin E, Emami J and Arabscheibani GR. (1998) A needs index for mental health care. *Social Psychiatry and Psychiatric Epidemiology,* 33, 89-96.

Goldberg DP and Huxley P. (1992) *Common mental disorders: a biosocial model*. Tavistock/Routledge, London.

Henderson S, Byrne DG and Duncan-Jones P. (1981) *Neurosis and social environment*. Academic Press, Sydney/New York.

Henderson S, Duncan-Jones P, Byrne DG, Scott R and Adcock S. (1979) Psychiatric disorder in Canberra. A standardised study of prevalence. *Acta Psychiatrica Scandinavica,* 60, 355-374.

Hennekens CH and Burning JE. (1987) *Epdemiology in medicine*. Lippincott Williams & Wilkins, Boston.

Hodiamont PPG, Peer N and Syben N. (1987) Epidemiological aspects of psychiatric disorder in a Dutch health area. *Psychological Medicine,* 17, 495-505.

Hodiamont PPG, Syben N, Koeter MW and Oldehinkel AJ. (1992) The relation of psychiatric disorder and urbanisation [in Dutch]. *Tijdschrift voor Psychiatrie,* 34, 426-438.

Hodiamont PPG and Veling SHJ. (1984) A model to assess psychiatric prevalence: The relation GHQ-PSE. [in Dutch]. *Tijdschrift voor Psychiatrie,* 26, 592-601.

Jenkins R, Lewis G, Bebbington P, Brugha T, Farrell M, Gill B and Meltzer H. (1997) The National Psychiatric Morbidity surveys of Great Britain-- initial findings from the household survey. *Psychological Medicine,* 27, 775-789.

Kessler RC, Anthony JC, Blazer DG, Bromet E, Eaton WW, Kendler K, Swartz M, Wittchen HU and Zhao S. (1997) The US National Comorbidity Survey: overview and future directions [editorial]. *Epidemiologia e Psichiatria Sociale,* 6, 4-16.

Kessler RC, Chiu WT, Demler O, Merikangas KR and Walters EE. (2005) Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry,* 62, 617-627.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU and Kendler KS. (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of General Psychiatry,* 51, 8-19.

Koeter MW and Ormel J. (1991) *General Health Questionnaire: Dutch version (in Dutch). Manual*. Swets and Zeitlinger, Lisse.

Kooij JJ, Buitelaar JK, van den Oord EJ, Furer JW, Rijnders CA and Hodiamont PP. (2005) Internal and external validity of attention-deficit hyperactivity disorder in a population-based sample of adults. *Psychological Medicine,* 35, 817-827.

Lesage AD, Cyr M and Toupin J. (1991) Reliable use of the Present State Examination by psychiatric nurses for clinical studies of psychotic and nonpsychotic patients. *Acta Psychiatrica Scandinavica,* 83, 121-124.

Lilienfeld AM and Lilienfeld DE. (1992) *Foundations of epidemiology*. Oxford University press, New York.

McConnell P, Bebbington P, McClelland R, Gillespie K and Houghton S. (2002) Prevalence of psychiatric disorder and the need for psychiatric care in Northern Ireland. Population study in the District of Derry. *British Journal of Psychiatry,* 181, 214-219.

Meertens V. (2004) *Depressive symptoms in the general population. A multifactorial social approach.* ICS-Thesis, Nijmegen.

Meertens V, Scheepers P and Tax B. (2003) Depressive symptoms in the Netherlands 1975-1996: a theoretical framework and an empirical analysis of socio-demographic characteristics, gender differences and changes over time. *Sociology of health and illness,* 25, 208-231.

Mootz M and van den Berg J. (1989) Indicators of health situations in the CBS-survey [in Dutch]. *Maandbericht Gezondheid,* 2, 4-10.

Narrow WE, Rae DS, Robins LN and Regier DA. (2002) Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 surveys' estimates. *Archives of General Psychiatry,* 59, 115-123.

Ormel J, Koeter MW, Van den Brink W and Giel R. (1989) Concurrent validity of GHQ-28 and PSE as measures of change. *Psychological Medicine,* 19, 1007-1013.

Peen J, Bijl RV, Spijker J and Dekker J. (2002) Does the prevalence rate of psychiatric disorders increase with urbanicity? [in Dutch]. *Tijdschrift voor Psychiatrie,* 44, 225-235.

Pincus HA, Tew JD and First MB. (2004) Psychiatric comorbidity: is more less? *World Psychiatry,* 3, 18-23.

Ravelli A, Bijl RV and van Zessen G. (1998) Comorbiditeit van psychiatrische stoornissen in de Nederlandse bevolking; resultaten van de Netherlands Mental Health Survey and Incidence Study (NEMESIS) [in Dutch]. *Tijdschrift voor Psychiatrie,* 40, 531-543.

Rijnders CATh and Furer JW. (2003) *Other people, different care: a study on determinants and consequences of comorbidity in the Nijmegen Health Area [in Dutch]* (Rep. No. 64496). Department of Social Medicine, RU, Nijmegen.

Rijnders CATh, van den Berg JF, Hodiamont PP, Nienhuis FJ, Furer JW, Mulder J and Giel R. (2000) Psychometric properties of the schedules for clinical assessment in neuropsychiatry (SCAN-2.1). *Social Psychiatry and Psychiatric Epidemiology,* 35, 348-352.

Robins LN, Locke BZ and Regier DA. (1991) An overview of psychiatric disorders in America. In LN Robins en DA Regier (Eds.), *Psychiatric disorders in America; The epidemiologic catchment area study*. pp. 328-366. The Free Press, New York.

Roca M, Gili M, Ferrer V, Bernardo M, Montano JJ, Salva JJ, Flores I and Leal S. (1999) Mental disorders on the island of Formentera: prevalence in general population using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Social Psychiatry and Psychiatric Epidemiology,* 34, 410-415.

Roscam Abbing EW and Hodiamont PPG. (2000) *High-Risk-Profile for psychopathology [in Dutch]* Department of Social Medicine, Nijmegen.

Susser E, Bresnahan M and Link B. (2002) Peering into the future of psychiatric epidemiology. In MT Tsuang en M Tohen (Eds.), *Textbook in psychiatric epidemiology, second edition*. pp. 195-211. Wiley-Liss, New York.

Thornicroft G and Tansella M. (2001) The planning process for mental health services. In G Thornicroft en G Szmukler (Eds.), *Textbook of community psychiatry*. Oxford University Press, Oxford.

Tomov T and Nikolov V. (1990) Reliability of SCAN categories and scores. Results of the field trials. In CN Stefanis (Ed.), *Psychiatry: a world perspective*. pp. 107-117. Elsevier Science Publishers, Amsterdam.

van den Akker M, Buntinx F and Knottnerus JA. (1996) Comorbidity or multimodality: what's in a name? A review of literature. *European Journal of General Practice,* 2, 65-70.

van den Berg J and van den Bos GAM. (1989) The (measurement of the) presence of chronic disorder, 1974 - 1987 [in Dutch]. *Maandbericht Gezondheid,* 3, 4-21.

van Rijswijk E. (2005) *Mental Health Problems in General Practice; an explorative study on diagnosis and treatment*. Thesis - Nijmegen University, Nijmegen.

van Rijswijk E, van de Lisdonk EH and Zitman FG. (2000a) Comorbidity and diagnosing depressive disorders in family practice. *Archives of Family Medicine,* 9, 123-124.

van Rijswijk E, van de Lisdonk EH and Zitman FG. (2000b) Who uses over-the-counter psychotropics? Characteristics, functioning, and (mental) health profile. *General Hospital Psychiatry,* 22, 236-241.

van Rijswijk E, Zandstra SM, van de Lisdonk EH, Zitman FG and van Weel C. (2005) Appropriateness of benzodiazepine prescribing in general practice. *International Journal of Clinical Pharmacology and Therapeutics,* 43, 411-412.

Vermeulen CA and Bosma AMM. (1992) *Health of the elderly; epidemiology and policy [in Dutch]*. Thesis UVA, Amsterdam.

WHO. (2000) Cross-national comparisons of the prevalences and correlates of mental disorders. WHO International Consortium in Psychiatric Epidemiology [In Process Citation]. *Bulletin of the World Health Organization,* 78, 413-426.

Wing JK. (1996) SCAN and the PSE tradition. *Social Psychiatry and Psychiatric Epidemiology,* 31, 50-54.

Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D and Sartorius N. (1990) SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry,* 47, 589-593.

Wing JK, Nixon JM, Mann SA and Leff JP. (1977) Reliability of the PSE (ninth edition) used in a population study. *Psychological Medicine,* 7, 505-516.

Wittchen HU, Üstün TB and Kessler RC. (1999) Diagnosing mental disorders in the community. A difference that matters? [editorial]. *Psychological Medicine,* 29, 1021-1027.

Zandstra SM, Furer JW, van de Lisdonk EH, Bor JH, Zitman FG and van Weel C. (2002a) Differences in health status between long-term and short-term benzodiazepine users. *British Journal of General Practice,* 52, 805-808.

Zandstra SM, Furer JW, van de Lisdonk EH, van't Hof HM, Bor JH, van Weel C and Zitman FG. (2002b) Different study criteria affect the prevalence of benzodiazepine use. *Social Psychiatry and Psychiatric Epidemiology,* 37, 139-144.

Zandstra SM, van Rijswijk E, Rijnders CA, van de Lisdonk EH, Bor JH, van Weel C and Zitman FG. (2004) Long-term benzodiazepine users in family practice: differences from short-term users in mental health, coping behaviour and psychological characteristics. *Family Practice,* 21, 266-269.

Part III

7 Psychometric properties of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN-2.1)

C.A.Th. Rijnders[1], J.F.M. van den Berg[1], P.P.G. Hodiamont[1,2], F.J. Nienhuis[3], J.W. Furer[1], J. Mulder[4], R. Giel[3]

[1] Dept. of Social Medicine, Nijmegen University, 117 SG, Post box 9101, 6500 HB Nijmegen, The Netherlands; [2] Dept. of Social Sciences, Tilburg University, The Netherlands; [3] Dept. of Social Psychiatry, Groningen University, The Netherlands; [4] Dept. of Data processing and Statistics, Nijmegen University, The Netherlands

Social Psychiatry and Psychiatric Epidemiology (2000), 35, 348-352

7.1 Introduction

Before any survey is started, insight into the psychometric properties of the proposed instrument is necessary to estimate the quality of the data that will be collected. This is even more pressing when the instrument is new or has been fundamentally revised.

The development of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN-2.1), one of the latest instruments developed by the WHO for the assessment of psychiatric disorders, is part of a rich tradition (Janca et al. 1994). The core part of the SCAN-2.1 is the tenth version of the Present State Examination (PSE-10), a semi-structured interview, and the successor to the PSE-9 (Wing et al. 1990; Wing 1996). Another instrument for the assessment of psychopathology is the Composite International Diagnostic Interview (CIDI), which is a highly structured interview, frequently used in epidemiological studies (Sartorius & Janca 1996). Interviewer-rated (semi-structured) interviews target the clinical situation better than highly structured interviews or self-rated questionnaires, probably at the cost of loss of reliability. Consequently it is essential to assess the psychometric properties of the SCAN-2.1. There is clear evidence of the reliability of the PSE-9 (Duine et al. 1985a; Wing et al. 1974) and also of the CIDI (Andrews & Peters 1998; Sartorius & Janca 1996; Wittchen et al. 1991; Wittchen 1994). Although some studies have been undertaken to obtain information on the reliability of specific sections of previous versions of the SCAN (Andrews et al. 1995; Easton et al. 1997; Tomov & Nikolov 1990; Wing et al. 1990), the psychometric properties of the entire interview SCAN-2.1 have never been tested.

In 1997 a general population survey was carried out in the Nijmegen Health Area (the Netherlands) with the Dutch version of the SCAN-2.1 (Giel & Nienhuis 1996) for psychiatric assessment. Prior to the survey data collection, a study on

the psychometric properties of the SCAN-2.1 had been carried out with the use of two designs, each with its own approach to the design-specific causes of measurement errors (Shrout 1995). In one design, pairs of independent live interviews with the same subject were compared (test-retest procedure). In the other, ten videotaped interviews by experts were rated by each of the interviewers (standardised procedure) and the outcomes were compared with the scores of the interviews conducted by the other interviewers and also with a reference score (the consensus score of the experts).

7.2 Subjects and methods

In addition to the PSE-10 semi-structured interview, the SCAN-2.1 includes two instruments recently developed by the WHO, the Clinical History Schedule and the Item Group Checklist, neither of which is an interview. They are used in clinical settings to obtain information from medical records or from significant others when the PSE-10 information from the subject is insufficient. These were not part of this study.
The PSE-10 consists of two parts: the first deals with anxiety, mood, and other neurotic signs and symptoms and the second with psychotic and cognitive signs and symptoms and observation items that target speech, behaviour and affect. For all data, specific DSM-IV diagnoses are computed by means of the SCAN-2.1 algorithm, which covers all the DSM-IV diagnoses except disorders usually first diagnosed in infancy, childhood or adolescence.

For the purpose of this study, specific DSM-IV diagnoses were assigned to three aggregation levels: closely related specific diagnoses (e.g. single-episode and recurrent depressive disorder) were classified into one diagnostic group (e.g. depressive disorder); associated diagnostic groups (e.g. depressive, dysthymic, and bipolar disorders) were classified into one diagnostic category (e.g. mood disorder), in accordance with the structure of the DSM-IV; and finally, the criterion for caseness was the presence of at least one specific diagnosis.
The same grouping procedure was used for syndromes, which were defined as complexes of psychopathologic symptoms, pursuant to the structure and description of the DSM-IV, no account having been taken of the criteria of organic attribution, duration, exclusion by other disorders or the explicit criteria of impaired functioning.
In this study, the entire interview from the SCAN-2.1 version published in July 1997 was tested, the final form of which contained no major changes.

Consistent with the clinical, semi-structured nature of the SCAN, 30 psychologists and psychiatric nurses were selected for their knowledge of psychopathology. Special sessions on psychiatric phenomenology were organised to optimize their actual clinical judgement. All interviewers participated in a 1-week training course given by the WHO-SCAN collaborating centre in Groningen so that they could learn how to work with the pencil-and-paper version of the SCAN. A total of ten booster sessions were held to improve their knowledge of psychopathology and their interviewing skills. The training consisted of reviewing videotaped interviews and discussing problems encountered during the data checks by the investigators. The interviewers were

also introduced to the computerised version of the SCAN, which had been approved by the WHO advisory committee for this instrument.

During the study, three interviewers dropped out. All the other interviewers (27) completed the study and subsequently took part in the survey data collection.

7.2.1 Test-retest procedure

Over a 3-month period in the spring of 1997, patients and non-patients were invited for an interview to be held twice in the same week. To guarantee a broad variety of disorders, the interviews were planned with three categories of subjects:

1 non (mental health care-identified) patients;
2 mental health care-identified out-patients;
3 mental health care-identified in-patients.

The non-patients were people with no known mental health care problem over the 3 months preceding the interview. They were invited to participate in the study in an announcement in a house-to-house circular or by their general practitioner during a visit. For the second and third categories (out-patients and in-patients), several mental health care institutions in the region were asked to invite their patients to participate in the study. For in-patients, this included Overwaal, a psychotherapy clinic, the Derde Orde Huis, an institute for psychosocial in-patient treatment, and the Psychiatric Centre Nijmegen, a psychiatric hospital. Out-patients were recruited from the RIAGG Nijmegen, an institute for ambulatory mental health care and the PCN out-patient department.

Every subject was interviewed twice in the same week by different, randomly assigned interviewers. In all interviews, only one interviewer was present. To facilitate supervision all interviews were videotaped. After completion of the second interview, the interview data were paired. Given the equivalence of the paired data, Cohen's kappa (κ) was considered an adequate measure of reliability. Kappa was estimated with regard to the three aggregation levels of diagnoses and syndromes (Landis & Koch 1977a; Landis & Koch 1977b). For a reliable κ, it is important to have a sufficient number of cases. Consequently, only diagnostic categories and diagnostic groups with more than ten cases were used.

7.2.2 Standardised procedure

Experts (RG, FJN and CAThR) interviewed three subjects with psychotic disorders, three with neurotic disorders and four non (mental health care-identified) patients. The video recordings of these SCAN interviews were rated by at least one other experienced psychiatrist (PPGH and CAThR). The data were combined and used as a reference score in the analysis of the results of the 27 interviewers, each of whom rated the ten videotapes. Because of the non-equivalence of the data of the interviewers and of the reference scores of the experts (gold standard), κ was not found to be the best parameter for agreement (Streiner & Geddis 1998). Consequently, sensitivity, specificity, and percentage of total agreement were assessed for the three aggregation levels of diagnoses and syndromes and also for the individual interviewers. Confidence intervals were computed with the jack-knife method.

7.3 Results

7.3.1 Test-retest procedure

During the study, 181 interviews were conducted with 92 subjects, three of whom were interviewed only once. This resulted in 89 paired interviews (Table 7.3.1).

Table 7.3.1
Number of subjects, interviews, and pairs of SCAN data per patient category

	Subjects	Interviews	Pairs of SCAN data
Non-patients	30	60	30
Out-patients	21	40	19
In-patients	41	81	40
Total	92	181	89

For diagnostic caseness, the κ was 0.64 (Table 7.3.2), which indicated that the reliability was substantial (Landis & Koch 1977b). With reference to the diagnostic categories and groups κ ranged from 0.24 (fair) to 0.64 (substantial).

Table 7.3.2
Test-retest reliability for diagnostic category, group and caseness on the basis of 89 sets of paired SCAN data

Category Group	Number of cases [a]	κ	95% CI
Mood disorder	26	0.53	0.32 - 0.75
Depression	23	0.52	0.29 - 0.75
Anxiety disorder	43	0.49	0.31 - 0.68
Phobic or panic disorder	16	0.48	0.21 - 0.76
Obsessive compulsive disorder	22	0.64	0.44 - 0.84
Anxiety disorder NOS [b]	12	0.24	0.00 - 0.55
Sleep disorder	12	0.35	0.01 - 0.68
Dyssomnia	11	0.38	0.04 - 0.72
Caseness	56	0.64	0.48 - 0.80

[a] If at least one of the interviewers scored a diagnosis, the subject was classified as a case; only groups and categories containing at least 10 cases were listed;

[b] not otherwise specified;

Table 7.3.3
Test-retest reliability for syndromes with reference to syndrome category, group and caseness on the basis of 89 sets of paired SCAN data

Category Group	Number of cases [a]	κ	95% CI
Mood syndrome	33	0.59	0.40 - 0.77
Depressive syndrome	27	0.59	0.39 - 0.80
Dysthymic syndrome	14	0.06	0.00 - 0.30
Anxiety syndrome	56	0.46	0.28 - 0.65
Phobic or panic syndrome	41	0.36	0.16 - 0.56
Obsessive compulsive syndrome	36	0.55	0.37 - 0.74
Somatoform syndrome	54	0.46	0.28 - 0.64
Somatoform syndrome NOS	54	0.46	0.28 - 0.64
Sleep syndrome	40	0.50	0.31 - 0.69
Parasomnia syndrome	35	0.54	0.35 - 0.73
Dyssomnia syndrome	26	0.49	0.27 - 0.71
Caseness	62	0.47	0.25 - 0.69

[a] If at least one of the interviewers scored a syndrome, the subject was classified as a case; only groups and categories containing at least 10 cases were listed

For syndrome caseness, κ was 0.47, which is slightly lower than for diagnosis. With reference to syndrome categories and groups, the κ-values were comparable to those found for diagnoses, except for the dysthymic syndrome (Table 7.3.3).

7.3.2 Standardised procedure

The reference scores from the ten videotaped interviews showed that three out of the four non-patients had been assessed as having had neither a psychiatric diagnosis nor a syndrome. The diagnosis of the fourth subject was a sexual disorder due to a general medical condition. Of the three subjects with psychotic disorders, two had a schizophrenic disorder (one of the two also had a diagnosis of a depressive disorder) and one had a delusional disorder combined with a bipolar disorder NOS. Of the three subjects with neurotic disorders, one had a depression combined with an obsessive compulsive disorder. The second subject was diagnosed as bipolar in the depressive phase, and the third had a diagnosis of cannabis dependence comorbid with a nightmare disorder.

For diagnostic caseness, the average total agreement of the 27 interviewers with the reference score was 94% (Table 7.3.4).

Table 7.3.4
Total agreement, sensitivity and specificity for diagnostic category, group and caseness.

Category Group	Total agreement %	Sensitivity %	Specificity %
Substance use-related disorder	97	67	100
Dependence	97	67	100
Psychotic disorder	87	56	100
Schizophrenia	81	9	99
Delusional disorder, non-aff. psychosis	93	89	93
Mood disorder	87	70	98
Depression	84	59	90
Bipolar disorder	90	50	100
Anxiety disorder	96	80	100
Phobic or panic disorder	95	52	100
Obsessive-compulsive disorder	100	100	100
Sleep disorder	97	93	97
Dyssomnia	98	93	98
Other psychiatric disorder	98	85	99
Caseness	94	86	99

The great majority of the cases were recognised by the interviewers (sensitivity: 86%) and almost no one assessed a case where the referents rated a non-case (specificity: 99%).

The most remarkable finding, however, was a sensitivity of only 9% to schizophrenic disorders, while in other diagnostic groups the sensitivity was in the 50-100% range. Sensitivity for diagnostic categories varied from 56% to 96%, whereas specificity ranged from 93% to 100% for both groups and categories.
The sensitivity and specificity for syndromes were excellent for syndrome caseness and also for categories and groups, with the exception of bipolar syndrome, for which sensitivity was 57% (Table 7.3.5). For schizophrenic syndrome (as opposed to schizophrenic disorder), sensitivity was as high as 79%.

Table 7.3.5
Total agreement, sensitivity and specificity for syndrome category, group and caseness.

Category Group	Total agreement %	Sensitivity %	Specificity %
Substance use-related syndrome	100	100	100
Dependence syndrome	100	100	100
Abuse syndrome	100	100	100
Psychotic syndrome	99	98	100
Schizophrenic syndrome	94	79	100
Delusional or non-affective psychotic syndrome	99	96	100
Mood syndrome	99	99	100
Depressive syndrome	98	96	100
Bipolar syndrome	91	57	100
Anxiety syndrome	96	80	100
Phobic or panic syndrome	96	80	100
Obsessive-compulsive syndrome	100	100	100
Somatoform syndrome	93	84	99
Pain syndrome	99	89	100
Other somatoform syndrome	93	84	99
Dissociation syndrome	99	93	100
Sleep syndrome	96	98	94
Parasomnia syndrome	96	98	94
Dyssomnia syndrome	97	96	97
Other psychiatric syndromes	97	96	97
Caseness	94	86	99

The interviewers were also judged on their individual agreement with the reference scores for each of the 10 videotaped interviews (Table 7.3.6). With regard to caseness, there was one reference score (case or non-case) for each interview, while for diagnostic categories, the number per tape varied from 1 to 3 (there were a total of 15 reference scores for the 10 tapes). For diagnostic groups, there were in total 14 reference scores. Since one interviewer scored substantially and consistently lower than all the others, Table 7.3.6 shows his scores separately.

The total mean agreement was 94% for diagnostic caseness, 90% for categories and 87% for groups. If the three tapes with the psychotic subjects had been left out in assessing reliability for part one of the SCAN, the average agreement (for the remaining seven tapes) would have been 98%, 96%, and 93%, respectively.

Table 7.3.6

Agreement per interviewer with regard to reference diagnoses per interview for the three diagnostic aggregation levels

	% Agreement		
	Diagnostic group N = 14	Diagnostic category N = 15	Caseness N = 10
Interviewer 1 - 26			
Range	82 - 95	84 - 100	90 - 100
Mean	87	91	95
Interviewer 27	70	73	74
Total mean	87	90	94

7.4 Discussion

Globally, the reliability of the SCAN-2.1 can be described as moderate to substantial. For the test-retest procedure, reliability was fair to moderate. The standardised procedure showed substantial to almost perfect agreement.

7.4.1 Study limitations

The lack of interview experience of the interviewers and instructions from the experts on the videotapes may have impacted negatively on the findings.

First, the most prominent exception in the standardised procedure, a low sensitivity for schizophrenia, can be explained as follows. To diagnose schizophrenia one has to record observations, the duration of the active symptoms and their interference with daily activities in four sections of the SCAN, in which there are no explicit questions. During the taped sessions, the interviewers were not reminded of the requirement to record these items, so the majority failed to rate them. This also explains the difference in sensitivity between the diagnosis of schizophrenia and its corresponding syndrome (9% and 79%), since, for the latter, duration and interference are not part of the definition of the pathology. Because the training was concentrated on the verbal part of the interview itself, we had to correct this imbalance during the booster sessions.

Second, although the limited number of cases for the dysthymic syndrome in the test-retest procedure reduced the significance of κ, the low reliability was probably caused by the confusing structure of the section in the SCAN on the depressed mood and thoughts. The diagnosis of dysthymic disorder depends completely on items that have explicit questions on depression in the beginning of this section. The answers corresponding to these items must be re-evaluated later on for their dysthymic quality without the aid of explicit questions.

Finally, a substantial number of the interviewers missed the bipolar diagnosis in one of the tapes even though they had correctly assessed a depressive episode. Although the interview data contained the necessary information, they failed to record two items (without explicit interview questions) that link the current depressive episode to the underlying bipolar disorder. This could be explained by the announcement by the expert on the tape that he would skip the rest of the section, whereas he actually recorded the two items himself.

All factors considered, lack of interviewing experience at the beginning of the study may have contributed to the above-mentioned lower reliability and agreement scores.

7.4.2 Literature

Although the literature contains no information on the psychometric properties of the SCAN-2.1, there is substantial evidence on the reliability of the SCAN-1, its predecessor the PSE-9, and the CIDI, a more structured interview for use in psychiatric epidemiological surveys. In the test-retest procedure adopted in this study, the entire SCAN-2.1 interview, with the full range of DSM-IV diagnoses, was used in a mixed population of subjects who reflected the psychiatric comprehensiveness of the general population. In contrast, the majority of the previous reliability studies referred to partial populations (clinical or outpatient) for which the diagnostic range was limited (e.g., substance-related or anxiety disorders) and to a restricted number of sections from a particular instrument (Andrews & Peters 1998; Duine et al. 1985a; Ustun et al. 1997; Wing et al. 1990).

In addition, it is noteworthy that the CIDI tends to produce more diagnoses than clinicians claim are clinically relevant (Andrews & Peters 1998); the SCAN-1 proved to be as reliable as the CIDI (Ustun et al. 1997); and lay interviewers can administer the SCAN reliably (Brugha et al. 1999b).
In view of all the preceding aspects, it can be concluded that the reliability of the SCAN-2.1 is satisfactory for the assessment of psychiatric disorders in general populations by non-clinical (but well-trained) interviewers. The findings from this study lead to the following recommendations:

During the 1-week training course, special attention should be paid to the structure of the SCAN and its underlying sections. The use of a checklist of items, which should be checked for their completeness at the end of the interview, might be very beneficial. Finally, from the standpoint of cost-effectiveness, it might be more efficient to train all interviewers intensively on part one of the SCAN, and to select a few of them for training on the entire SCAN interview on the basis of a superior knowledge of psychopathology.

References

Andrews G and Peters L. (1998) The psychometric properties of the Composite International Diagnostic Interview. *Social Psychiatry and Psychiatric Epidemiology,* 33, 80-88.

Andrews G, Peters L, Guzman AM and Bird K. (1995) A comparison of two structured diagnostic interviews: CIDI and SCAN. *Australian and New Zealand Journal of Psychiatry,* 29, 124-132.

Brugha TS, Nienhuis F, Bagchi D, Smith J and Meltzer H. (1999) The survey form of SCAN: the feasibility of using experienced lay survey interviewers to administer a semi-structured systematic clinical assessment of psychotic and non-psychotic disorders. *Psychological Medicine,* 29, 703-711.

Duine TJ, Hodiamont PPG, Peer PGM and Boerma LH. (1985) Betrouwbaarheidsaspecten van de Present State Examination in vier verschillende populaties [in Dutch]. *Tijdschrift voor Psychiatrie,* 27, 326-340.

Easton C, Meza E, Mager D, Ulug B, Kilic C, Gogus A and Babor TF. (1997) Test-retest reliability of the alcohol and drug use disorder sections of the schedules for clinical assessment in neuropsychiatry (SCAN). *Drug and Alcohol Dependence,* 47, 187-194.

Giel R and Nienhuis FJ. (1996) *SCAN-2.1: Schedules for clinical assessment in neuropsychiatry [in Dutch]*. WHO, Geneva / Amsterdam.

Janca A, Üstün TB and Sartorius N. (1994) New versions of World Health Organization instruments for the assessment of mental disorders. *Acta Psychiatrica Scandinavica,* 90, 73-83.

Landis JR and Koch GG. (1977a) An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics,* 33, 363-374.

Landis JR and Koch GG. (1977b) The measurement of observer agreement for categorical data. *Biometrics,* 33, 159-174.

Sartorius N and Janca A. (1996) Psychiatric assessment instruments developed by the World Health Organization. *Social Psychiatry and Psychiatric Epidemiology,* 31, 55-69.

Shrout PE. (1995) Reliability. In Tsuang, Tohen en Zahner (Eds.), *Textbook in Psychiatric Epidemiology*. pp. 213-227. Wiley-Liss, inc..

Streiner D and Geddis J. (1998) Some useful concepts and terms used in articles about diagnosis. *Evidence-based Mental Health,* 1, 7.

Tomov T and Nikolov V. (1990) Reliability of SCAN categories and scores. Results of the field trials. In CN Stefanis (Ed.), *Psychiatry: a world perspective*. pp. 107-117. Elsevier Science Publishers, Amsterdam.

Ustun B, Compton W, Mager D, Babor T, Baiyewu O, Chatterji S, Cottler L, Gogus A, Mavreas V, Peters L, Pull C, Saunders J, Smeets R, Stipec MR, Vrasti R, Hasin D, Room R, Van den Brink W, Regier D, Blaine J, Grant BF and Sartorius N. (1997) WHO Study on the reliability and validity of the alcohol and drug use disorder instruments: overview of methods and results [published erratum appears in 1998 Apr 1;50(2):185-6]. *Drug and Alcohol Dependence,* 47, 161-169.

Wing JK. (1996) SCAN and the PSE tradition. *Social Psychiatry and Psychiatric Epidemiology,* 31, 50-54.

Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D and Sartorius N. (1990) SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry,* 47, 589-593.

Wing JK, Cooper JE and Sartorius N. (1974) *The desciption and classification of psychiatric symptoms. An instruction manual for the PSE and CATEGO system*. Cambridge University Press, Cambridge.

Wittchen HU. (1994) Reliability and validity studies of the WHO--Composite International Diagnostic Interview (CIDI): a critical review. *Journal of Psychiatric Research,* 28, 57-84.

Wittchen HU, Robins LN, Cottler LB, Sartorius N, Burke JD and Regier D. (1991) Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI). The Multicentre WHO/ADAMHA Field Trials [published erratum appears in Br J Psychiatry 1992 Jan;160:136]. *British Journal of Psychiatry,* 159:645-53, 658, 645-53, 658.

8 The relevance of clinical judgement in the SCAN-2.1

8.1 Introduction

The essential questions in (descriptive) epidemiology are: "What condition occurs in whom, where and when?" (see chapter 3). It is the psychiatric diagnostic process that makes or breaks the quality of the assessment of the condition. In the past half century, a lot of energy has been spent on improving the quality of the diagnostic process. In the mid-1950s, under the influence of developments in biological, pharmacological and epidemiological research, the focus of the diagnostic process shifted from an individualised or idiographic to a categorical or nomothetic approach. Evidence from a global perspective indicated that there was a lack of uniformity with regard to the psychiatric diagnostic process. The US-UK study and the International Pilot Study of Schizophrenia (IPSS) established that the diagnostic discrepancy did not reflect a true difference in psychiatric prevalence rates (Kendell et al. 1971; Kendell 1975a; Sartorius et al. 1972), but rather a diagnostic inaccuracy, due to three methodological sources of variance:

1. the information variance;
2. the observation and interpretation variance;
3. the criterion variance.

The information variance concerns the process of gathering information, which relates to, for example, the kind and order of the questions and the context in which the interview is held. The observation and interpretation variance has to do with inconsistencies in interpreting the meaning of the explicit and circumstantial information. Finally, the criterion variance is about the variation in the reference criteria against which the information is interpreted. Ward et al. (1962b) elaborated on the three basic sources of variance and, at the time, regarded the criterion variance, which accounted for two-thirds of the total variance, as the most important of the three, while the former two, were estimated to make a joint contribution of only one third of the total variance.

Because the US-UK study and the IPSS stressed the importance of the criterion variance, it is this source of variance that helped drive the development of the current criterion-based psychiatric classification systems. Nowadays, the criterion variance is significantly less important, because of the almost global use of standard classification systems.

With a view to minimising the remainder of the variance, two lines of standardised diagnostic interviews were developed: the fully standardised interviews of which the DIS/CIDI[viii] tradition is the most well-known and the semi standardised interviews, including the PSE/SCAN[ix]. With respect to the

information variance, the two interview traditions differ in that the DIS/CIDI is fully structured (with the requirement to put the designated questions only) and the PSE/SCAN is semi-structured (with the recommendation to probe further until the interviewer is able to score the item). With respect to the observation and interpretation variance, the two traditions use divergent approaches. The DIS/CIDI has the response of the subject for each sign and symptom as the sole source of information, sometimes followed by a set of questions to check for clinical significance. This results in a measure of perceived morbidity or illness. The PSE/SCAN facilitates the interviewer's clinical judgement, bringing about an opportunity for such judgement for each sign and symptom. This results in a measurement of disease, in the form of a characteristic constellation of signs and symptoms, as diagnosed on the basis of clinical (professional) judgement (see chapter 3).
Throughout several decades, the respective interview traditions have yielded different prevalence rates (Brugha et al. 2001; Narrow et al. 2002; Surtees et al. 1983). The impact of clinical judgement and of the clinical significance criterion on this variation is still unclear (Beals et al. 2004; Murphy 2002).

To make clinical judgement operational, it is important to identify the aspects of signs and symptoms that should be clinically assessed. So that a sign or symptom can be evaluated as being clinically relevant, in version 2.1 of the SCAN the following questions have to be addressed qualitatively:

- Does the symptom produce sufficient distress?
- Does the symptom occupy a disproportional place the subject's life?
- Does the symptom escape the subject's control?

Having been made operational in this way, clinical judgement was taken into account in the rating scale of the SCAN-2.1. Not only can a symptom be rated as being clinically significant as indicated by score 2 (moderate) and 3 (severe); it can also be rated as present, but at a subclinical level (score 1). Hence, the symptom is recorded, but has no value in meeting the criteria for a disease.

Both positive and negative answers by the subject can result in a (clinical or subclinical) positive or negative score by the observer (Table 8.1.1). In other words, perceived morbidity (illness, response "yes") can indicate that a symptom is present, but without the formal value of a disease entity (subclinical). In contrast, in a fully structured interview, the subject's answers are recorded in accordance with a presence-absence dichotomy.

Table 8.1.1
Combination of illness and disease scores with respect to signs and symptoms

Response	No	No	No	No	Yes	Yes	Yes	Yes
Symptom	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Quality		Sub	Moderate	Severe		Sub	Moderate	Severe
Symptom status		clinical				clinical		
Illness score	0	0	0	0	1	1	1	1
Disease score	0	1	2	3	0	1	2	3
Illness 'result'	-	-	-	-	+	+	+	+
Disease 'result'	-	-	+	+	-	-	+	+

The difference between the two interview traditions resulting from the clinical judgement of the symptoms is indicated in the pale-shaded area of Table 8.1.1. The difference in diagnostic status of the symptoms in the two interviews is shown in the grey- and black-shaded areas of Table 8.1.1.

The SCAN-2.1 is a semi-structured, clinical diagnostic interview that creates the opportunity to elaborate on the effect of the clinical judgement on the diagnostic status, in particular with respect to the subclinical score (1-1), for which the result is represented by the black area of Table 8.1.1. The other subclinical score (0-1), which in itself is indistinguishable from the former score, is much less important for the following reason:

Except in psychotic disorders, the likelihood that a subject's negative answer will result in a positive (sub-)clinical score is very low. The same applies to the first two left-hand cells of the grey-shaded area, with a dissimilar result for the measurement of illness and disease, a difference (although the other way around) that also occurs when a persistent positive response is rated as totally absent (third left-hand cell). For this reason, we assume that the importance of the clinical judgement in the SCAN is best reflected by the use of the subclinical score.

The following questions constitute a guideline on the importance of the clinical judgement with the SCAN.

- Does the clinical judgement in the SCAN make a difference?
  - What is the number of positive cases and of diagnostic groups when the SCAN is used in accordance with the formal protocol and when clinical judgement is omitted?
- Does the omission of the clinical judgement have consequences for the epidemiological results?
  - What is the prevalence rate of psychiatric disorders across caseness, categories and groups with and without clinical judgement?
  - Does clinical judgement influence the distribution of psychiatric disorders across gender, age, and urbanisation?
  - What is the influence of clinical judgement on the comorbidity rate?

8.2 Methods

8.2.1 Design of the basic study

In 1997 the Nijmegen Health Area-2 project (NHA-2) was carried out in the general population of the city of Nijmegen and the surrounding communities (SC). In a sample of 1813 Dutch-speaking subjects aged from 18 to 75 years, psychiatric morbidity was assessed in a two-phase design. In phase one (T1) the General Health Questionnaire (GHQ-30) was used for screening, while in phase two (T2, see chapter 6) the psychiatric signs and symptoms were assessed with the Schedules for Clinical Assessment in Neuropsychiatry (SCAN-2.1). In this way, at T2, detailed information with reference to individual symptoms was obtained from 767 subjects, which created the opportunity to investigate the consequences of the application of clinical judgement in the SCAN.

8.2.2 Operationalising clinical judgement in SCAN

In T2, the subjects were interviewed with the SCAN-2.1, which includes the latest (tenth) version of the PSE (PSE-10), the successor to PSE-9 (Wing et al. 1990; Wing 1996). The PSE-10 consists of two parts, the first dealing with anxiety, mood and other neurotic manifestations and the second with psychotic and cognitive phenomena, and observation items on speech, behaviour and affect. In this survey, the interview in the July 1997 version of SCAN-2.1 was used (Giel & Nienhuis 1997). As this was a preliminary version of the current SCAN-2.1, the software algorithm was written by our own staff (C.R.) in consultation with, and subject to the approval of, the WHO workgroup on the SCAN. This facilitated the realisation of the modifications of the algorithm necessary for this study.

Consistent with the clinical, semi-structured nature of the SCAN, 30 psychologists and psychiatric nurses were selected for their knowledge of psychopathology. Special sessions on psychiatric phenomenology were organised to optimise their clinical judgement. All interviewers participated in a 1-week training course given by the WHO-SCAN collaborating centre in Groningen so that they could learn to work with the pencil-and-paper version of the SCAN. A total of ten booster sessions were held to improve their knowledge of psychopathology and their interviewing skills.

The claim that clinical interviews are imperfect because of low interrater reliability and potential validity problems (Wittchen et al. 1999) has been rebutted for the Present State Examination-9 (PSE-9: (Duine et al. 1985b; Lesage et al. 1991; Wing et al. 1977)) and for the Schedules for Clinical

Assessment in Neuropsychiatry-2.1 (SCAN-2.1: (Andrews et al. 1995; Brugha et al. 1999b; Cheng et al. 2001; Rijnders et al. 2000; Tomov & Nikolov 1990)).
Clinical judgement in the SCAN has been incorporated into pre-defined rating scales with the following basic format (Giel & Nienhuis 1997):

0. A positive rating of absence. It does not mean 'not known' or 'uncertain whether present or not'[x]. It can be used only if sufficient information is available to establish its accuracy.
1. A positive rating of presence, but presence in such a slight degree that it is not appropriate for use in classification. Like 0, it does not mean 'not known' or 'uncertain'.
2. The item is present at a level sufficient to use in classification. For this purpose it is equivalent to 3. In general, it is used when symptoms are of moderate severity during most of the period being assessed.
3. Similar to 2, except that the symptom is present in severe form for most of the period under review.

The formally used algorithm of SCAN-2.1 uses level 2 and 3 symptoms to assess disorders in accordance with DSM-IV. In this chapter, we have treated the subclinical scores corresponding to a rating of 1 as being equal to the clinical scores 2 and 3, resulting in the inclusion of all perceived morbidity.

8.2.3 Analyses

For all analyses, the Statistical Analyses for Social Sciences software package (SAS v8.2) was used.
In view of the evaluation of the clinical judgement in the SCAN, the NHA-2 data were re-analysed, essentially with the same algorithm as in chapter 6, but in this case, for each symptom for which it was necessary to assess whether specific DSM-criteria were fulfilled, the subclinical score (1) was assigned a weight equal to that of the clinical score (2). For the purpose of this survey, SCAN diagnoses were grouped on three levels: first, closely related diagnoses like single episode and recurrent depressive disorder were sorted into one *diagnostic group* (e.g., depressive disorder); then, associated diagnostic groups (e.g., depressive, dysthymic, and bipolar disorders) were classified into one *diagnostic category* (e.g., mood disorder), in accordance with the structure of DSM-IV; finally, the highest level of aggregation was represented by *diagnostic caseness,* defined as "the presence of at least one specific diagnosis".
The relation between the GHQ and SCAN "caseness" was expressed in a logistic regression model in which the probability of an individual's qualifying for "caseness" was formulated as a linear function of the GHQ score:

$$p(\text{caseness}) = \frac{1}{1+e^{-(\alpha + \beta * GHQ)}}$$

where p indicates the probability that a certain GHQ score corresponds to SCAN caseness on the basis of the assessed α and β,which reflect the relationship between the GHQ and the SCAN caseness as estimated from the outcome of all T2 interviews (Henderson et al. 1979; Hodiamont & Veling 1984). Because only 25% of the subjects with a GHQ score of less than 4 during their T1-interview were interviewed in T2, a weighting procedure was used for this group to restore the proportions initially present in the T1 group. Through substitution of the assessed probability for SCAN caseness for the GHQ score for all subjects in T1, the prevalence rate for the population was calculated.

The prevalence rates for the diagnostic groups and categories were calculated through weighing the extent to which they occurred in the two GHQ classes with scores < 4 and ≥ 4 for T2 subjects. The weighting factors were determined by the ratio of the sizes of the two GHQ classes to one another in the first phase of the sample. The significance of the difference between results with and without clinical judgement was tested with the Wilcoxon (Mann-Whitney) procedure.

So that the data for the general population could be interpreted, a post-stratification weighting procedure was applied with the aid of 16 defined strata for gender, age (4), and urbanisation (2) on the basis of the 1997 census data (CBS 1997).

To estimate the 95% two-sided confidence intervals for these weighted data, the logistic model was integrated into a bootstrap procedure, i.e. a re-sampling technique with duplication (Dunn et al. 1999; Efron & Tibshirani 1993).

The significance of the difference in prevalence rates within sociodemographic distributions was tested in an ANOVA model.

8.3 Results

8.3.1 The difference between results obtained with and without clinical judgement

In the NHA-2 study, 767 subjects were interviewed with the SCAN-2.1, 197 of whom fulfilled the DSM-IV criteria of psychiatric caseness. When clinical judgement was omitted from the formal SCAN procedure, the number of positive cases increased by more than a fifth, to 245 subjects (Table 8.3.1).

With reference to the diagnostic categories, the most substantial differences in numbers of positive cases between results obtained with and without clinical judgement were found in somatoform and psychotic disorders (almost a 100% increase) and mood disorders (nearly a 50% increase). When clinical judgement was omitted, anxiety disorders increased by around 20%. No change was found for substance-related, dissociation and eating disorders. An important similarity of the latter categories is that the majority of their diagnostic criteria refer to aspects of behaviour that leave (almost) no possibility for subclinical scores or for feelings and beliefs.

For the somatoform category, all additional diagnoses concerned the diagnostic group of somatoform disorders not otherwise specified (NOS) – a catch basin of single, unexplained, physically perceived conditions. Most subjects in question had a subclinical score on the single item of perceived fatigability and exhaustion, which changed their status as a healthy subject into a case. Only four of the subjects also had two to four somatic symptoms. These could be attributed to somatisation, but were not sufficient to qualify as a somatisation disorder.

Of the 25 additional cases in the category of mood disorders, 23 subjects were assessed as having a depressive episode, one subject a combined depressive and dysthymic disorder and one subject a bipolar disorder. Table 8.3.2 shows that the effect of clinical judgement on the number of DSM-IV criterion A symptoms for depression was substantial (a classical example of possible divergence between perceived morbidity and clinical relevance). For those subjects with an additional diagnosis of depression, the increase in the number of scored criterion A symptoms when clinical judgement was omitted ranged from 1 (for four subjects) to 4 or more (for seven subjects).

The ten subjects with an additional anxiety disorder presented a more varied picture: four had a generalised anxiety disorder, four had a phobic disorder and two had an anxiety disorder NOS.

As for sleep disorders, the eight additional subjects all had dyssomnia.

Remarkably, in addition to the near doubling of subjects with a somatoform disorder, three subjects lost their diagnosis when clinical judgement was omitted.

Table 8.3.1
Comparison of number of positive cases and diagnostic categories with and without clinical judgement (N = 767)

	With clinical judgement	Without clinical judgement	
	Cases / diagnoses	Additional positive cases / diagnoses	Lost positive cases / diagnoses
Caseness	197	48	--
Psycho-organic disorder	2	--	--
Substance-related	41	--	--
Psychotic disorder	4	3	--
Mood disorder	54	25	--
Anxiety disorder	54	10	--
Eating disorder	5	--	--
Somatoform disorder	31	27	3
Dissociation disorder	14	--	--
Sleep disorder	48	8	9
Other psychiatric disorder	7	4	--

The addition of eight diagnoses with reference to sleep disorders was more than neutralised by the loss of subjects who had a formal sleep disorder, but who now lost their sleep disorder diagnosis due to the omission of clinical judgement. This did not mean that subjects who had 'lost' their specific sleep disorder diagnosis would no longer be a case; all of those subjects who had lost their sleep disorder diagnosis had dyssomnia, which turned into a part of a diagnosis of a higher hierarchy (for example depressive disorder) when clinical judgement was left out. Parasomnia was consequently not related to the variation in the category of sleep disorders.

Table 8.3.2
Number of criterion A symptoms for depressive episode in subjects with an additional diagnosis of depression when clinical judgement was left out

		With clinical judgement				
	Criterion A symptoms	0	1	2	3	4
Without clinical judgement	5	--	1	2	2	4
	6	1	--	2	3	5
	7	--	--	1	1	1
	9	--	--	--	--	1

The contribution of the additional diagnoses listed in Table 8.3.1 to monomorbidity and to comorbidity is shown in Table 8.3.3. Psychotic and

anxiety disorders, unlike all other diagnoses, did not contribute more to monomorbidity than to comorbidity. Almost all additional diagnoses contributing to monomorbidity were 'new' cases, the vast majority having been (minor) somatoform and sleep disorders. Four subjects were already monomorbid and one subject was comorbid when the formal SCAN algorithm was used but became monomorbid when clinical judgement was omitted. The diagnoses for the latter five subjects were transposed into a depressive disorder, which, owing to the hierarchy, switched off the underlying formal diagnoses.

Additional disorders contributing to comorbidity mainly had the effect of changing monomorbidity into comorbidity; to a lesser degree these disorders became part of an already existing comorbidity. Rarely did they result in the attribution of a comorbid state to healthy subjects.

Consequently, within the group of subjects who constituted a case originally, omitting clinical judgement also had a qualitative effect with respect to the shift of specific diagnosis.

Although many diagnostic groups showed an increase when clinical judgement was omitted, there was a marked difference in this increase between groups. Of the 25 new mood disorders diagnosed when clinical judgement was omitted, 13 were new cases, while 12 represented a qualitative shift of the specific diagnosis. Six out of 10 and 2 out of 3 new diagnoses for anxiety disorders and psychotic disorders respectively resulted from the qualitative shift. In contrast, 22 of the 27 new somatoform disorders and 7 out of the 8 new sleep disorders were new cases.

Table 8.3.3

Contribution of the additional diagnoses to monomorbidity and comorbidity with respect to caseness and diagnostic categories

	Caseness	Psychotic disorders	Mood disorders	Anxiety disorders	Somatoform disorders	Sleep disorders	Other psychiatric disorders
Non → Mono*	46	1	12	3	21	7	2
Mono → Mono	--	--	4	--	--	--	--
Como → Mono	--	--	1	--	--	--	--
Non → Como	2**	--	1	1	1	--	--
Mono → Como	--	--	5	3	5	1	2
Como → Como	--	2	2	3	--	--	--
Total	48	3	25	10	27	8	4

* Non = Non-morbidity
Mono = Monomorbidity
Como = Comorbidity
The first condition resulted from the formal SCAN and the second from the omission of clinical judgement.

** One comorbid subject had two mood disorders (a depression and a dysthymic disorder).

8.3.2 The consequences of the omission of clinical judgement on epidemiological data

8.3.2.1 Clinical judgement and the prevalence rate of psychiatric disorders

The effect of the omission of clinical judgement on the prevalence rates of caseness, diagnostic categories and diagnostic groups in the general population is presented in Table 8.3.4.

The prevalence rate of caseness without clinical judgement was 21.4%, which is 3.8% higher than when the formal clinical SCAN-algorithm was used. In other words, caseness increased by one fifth when clinical judgement was not used.

The prevalence rates of somatoform, mood and anxiety disorders were significantly higher when clinical judgement was not used. The prevalence rates of other psychiatric disorders, sleep disorders and psychotic disorders tended to increase and the prevalence rates of psycho-organic, substance-related, dissociation and eating disorders were stable and consequently seem to be independent of clinical judgement.

With respect to the diagnostic groups, the prevalence rate of other somatoform disorders more than doubled, and for depression it nearly doubled. The prevalence rate of the category ‘other psychiatric disorders’ doubled, but owing to the smallness of the sample size in the category, it could not be established whether the differences were statistically significant.

Table 8.3.4

Prevalence rates[a] of psychiatric disorders with and without clinical judgement with reference to caseness, diagnostic categories and groups

	Prevalence (%)		p
	With clinical judgement	Without clinical judgement	
Caseness	17.6 (± 2.6)	21.4 (± 3.0)	***
Psycho-organic disorder	0.3	0.3	
Dementia	0.3	0.3	
Other organic brain disorder	0.0	0.0	
Substance-related disorder	4.0	4.0	
Dependence	0.9	0.9	
Abuse	3.2	3.2	
Psychotic disorder	0.5	0.7	
Schizophrenia	0.4	0.5	
Delusional disorder, non-aff. psychosis	0.1	0.2	
Mood disorder	3.8	5.6	***
Depression	2.7	4.5	***
Mania, hypomania, mixed disorder	0.3	0.4	
Dysthymic disorder	1.9	2.0	
Anxiety disorder	4.6	5.5	**
Generalised anxiety disorder	0.3	0.6	
Phobic or panic disorder	2.6	3.0	
Obsessive-compulsive disorder	1.3	1.3	
Other anxiety disorders	0.6	0.7	
Eating disorder	0.4	0.4	
Anorexia Nervosa	0.1	0.1	
Other eating disorders	0.3	0.3	
Somatoform disorder	2.8	4.9	***
Somatisation disorder	0.1	0.1	
Pain disorder	1.2	1.0	
Other somatoform disorder	1.5	3.7	***
Dissociation disorder	1.4	1.4	
Sleep disorder	4.8	5.1	
Dyssomnia	2.2	2.2	
Parasomnia	2.1	2.1	
Other sleep disorder	0.5	0.8	
Other psychiatric disorder	0.7	1.4	

[a] after correction for gender, age, and urbanisation

p: Mann-Whitney: *: .01 < p ≤ .05; **: .001 < p ≤ .01; ***: p ≤ .001

The total increase in the diagnostic categories was 6.0%, which was strikingly higher than the 3.8% increase for caseness. This discrepancy was mainly a consequence of the increase in comorbidity, and consequently of the qualitative shift of specific diagnoses.

8.3.2.2 Clinical judgement and the distribution of psychiatric disorders

When clinical judgement was omitted, the sociodemographic groups with the highest and those with the lowest prevalence rates all showed roughly the same proportional increase in caseness (Table 8.3.5). As a result, the sociodemographic groups with the highest prevalence rates increased the most numerically. Likewise, the lowest prevalence rates increased the least numerically. It can be concluded that generally, clinical judgement does not affect the distribution of psychiatric disorders in the general population.

For those diagnostic categories and groups for which clinical judgement affected the prevalence rate (Table 6.3.3), the sociodemographic distribution across gender, age and urbanisation is presented with reference to clinical judgement in Table 8.3.5.

Globally, clinical judgement for diagnostic categories brought about larger proportional changes for females than for males, in particular for somatoform (2.04 vs. 1.39), anxiety (1.25 vs. 1.03) and other psychiatric disorders (4.00 vs. 1.00, but with low numbers), while for mood disorders, for instance, males were more affected (1.35 vs. 1.81). The underlying diagnostic groups showed the same trends.

With respect to age, the effect of clinical judgement varied for most of the diagnostic categories. The largest proportional increase for somatoform disorders occurred in the younger age range (≤ 44 yrs, 2.22). As far as sleep disorders were concerned, the older subjects (≥60 yrs) were more affected (1.34), especially in the case of dyssomnia, while the youngest age group actually showed a decrease (0.78). For mood and anxiety disorders, the middle-aged subjects (aged 45 - 59 yrs) were relatively more affected (1.58 and 1.38), but the highest proportional increase for mood disorders occurred in the youngest age group (1.78).

City dwellers had a slightly larger proportional increase in anxiety disorders (1.32 vs. 1.16), while rural dwellers tended to present a larger effect on mood disorders (1.59 vs. 1.30) when clinical judgement was omitted.

Table 8.3.5
Sociodemographic distribution of psychiatric caseness with and without clinical judgement (Males)

	Total			Males					
	N = 1797			Nijmegen N = 214			SC N = 589		
Clinical judgement		Yes	No		Yes	No		Yes	No
	N	%	%	N	%	%	N	%	%
Caseness	1799	17.6	21.4	214	18.8	22.6	589	14.8	18.1
Age group (yrs)									
18-24	120	16.0	19.7	16	21.9	26.3	28	12.4	15.5
25-34	359	18.4	22.4	44	18.5	22.8	93	13.2	16.2
35-44	461	16.8	20.4	64	17.9	21.6	151	15.2	18.5
45-54 [a]	391	20.1	25.2	41	20.7	24.8	144	18.1	21.8
55-64	272	16.7	20.2	26	19.8	23.1	102	15.2	18.4
≥ 65 [b]	194	14.7	17.8	23	14.9	17.9	70	12.7	15.3
Marital status									
Single	141	19.8	24.4	41	22.2	27.2	33	20.8	25.4
Living with parents	68	15.4	19.0	5	24.5	29.0	32	13.2	16.3
Living with partner	214	18.2	22.1	35	17.9	21.8	39	17.0	20.5
Married	1209	17.2	20.8	117	18.1	21.5	454	14.4	17.5
Divorced	88	22.7	27.2	11	15.2	18.9	24	20.6	24.4
Widowed	77	16.9	20.7	5	14.2	17.8	6	19.3	24.1
Employment status **									
Full-time [bd]	706	15.6	19.1	130	16.2	19.9	384	14.4	17.5
Part-time [bd]	378	18.0	21.9	23	15.1	18.5	27	14.3	17.2
Retired [bd]	189	14.8	17.7	28	15.6	18.8	100	14.7	17.6
Chronically ill/unable to work [a]	74	29.5	34.7	6	58.8	65.7	42	21.5	26.0
Unemployed [c]	40	29.5	34.6	10	35.0	39.8	12	20.3	24.5
School or college student [b]	60	17.8	21.8	12	24.6	29.3	14	15.4	19.3
Householder	285	19.9	24.0	--	--	--	1	7.5	8.7
Others without an occupation [b]	59	17.4	20.8	5	25.7	29.7	7	16.4	20.3
Educational level									
University	151	17.4	21.3	46	18.7	22.6	35	16.0	19.5
Higher vocational education	314	15.3	18.7	36	16.8	20.7	97	14.0	17.1
Higher secondary education	162	17.9	21.7	36	19.5	23.5	40	10.9	13.3
Lower-stream sec. & post-sec. lower vocational education	338	16.0	19.4	26	17.2	21.3	130	14.6	17.6
Lower-stream sec. education	226	18.5	22.3	18	17.9	21.2	55	16.7	20.6
Continued primary education	423	19.7	23.6	38	21.9	25.5	156	15.7	18.9
No education or only primary education	183	20.0	24.2	14	17.9	21.0	75	17.5	21.2

** N = 6: missing

Age group: Kruskal-Wallis: .0001 < p < .001; Scheffé, p = .05 [a] – [b]

Marital status: Kruskal-Wallis: .001 < p < .01; Scheffé, not significant

Employment status: Kruskal-Wallis: p < .0001; Scheffé, p = .05 [a] – [b], [c] – [d]

Educational level: Kruskal-Wallis: .001 < p < .01; Scheffé, not significant

Table 8.3.6 (continued)
Sociodemographic distribution of psychiatric caseness with and without clinical judgement (Females)

	Total			Females					
				Nijmegen			SC		
	N = 1797			N = 311			N = 685		
Clinical judgement		Yes	No		Yes	No		Yes	No
	N	%	%	N	%	%	N	%	%
Caseness	1799	17.6	21.4	311	21.3	25.7	685	17.8	21.7
Age group (yrs)									
18-24	120	16.0	19.7	29	19.6	24.0	47	13.9	17.3
25-34	359	18.4	22.4	86	20.9	25.3	136	20.4	24.7
35-44	461	16.8	20.4	80	19.8	23.8	166	16.5	20.1
45-54 [a]	391	20.1	25.2	53	25.3	30.4	153	20.0	26.6
55-64	272	16.7	20.2	28	26.9	31.8	116	14.9	18.3
≥ 65 [b]	194	14.7	17.8	35	17.6	21.2	66	15.2	18.5
Marital status									
Single	141	19.8	24.4	48	19.1	23.6	19	14.9	18.5
Living with parents	68	15.4	19.0	3	25.0	30.1	28	15.4	19.0
Living with partner	214	18.2	22.1	84	20.1	24.2	56	16.5	20.4
Married	1209	17.2	20.8	124	22.6	27.0	514	18.2	22.1
Divorced	88	22.7	27.2	31	27.3	32.7	22	22.2	26.7
Widowed	77	16.9	20.7	21	16.3	19.5	45	17.2	21.2
Employment status **									
Full-time [bd]	706	15.6	19.1	92	16.2	20.0	100	18.7	23.0
Part-time [bd]	378	18.0	21.9	90	21.6	26.2	238	17.4	21.2
Retired [bd]	189	14.8	17.7	20	20.8	24.7	41	11.3	13.8
Chronically ill/unable to work [ae]	74	29.5	34.7	9	36.3	40.3	17	35.4	42.6
Unemployed [c]	40	29.5	34.6	9	39.5	46.9	9	25.6	30.0
School or college student [b]	60	17.8	21.8	13	19.3	23.6	21	14.6	18.1
Householder [f]	285	19.9	24.0	57	26.9	31.9	227	18.2	22.1
Others without an occupation [b]	59	17.4	20.8	19	16.0	19.0	28	17.1	20.6
Educational level									
University	151	17.4	21.3	62	17.1	21.0	8	18.4	23.1
Higher vocational education	314	15.3	18.7	82	16.8	20.6	99	14.7	18.0
Higher secondary education	162	17.9	21.7	35	21.7	26.6	51	19.7	23.9
Lower-stream sec. & post-sec. lower vocational education	338	16.0	19.4	16	24.4	28.3	166	16.1	19.6
Lower-stream sec. education	226	18.5	22.3	32	20.5	24.4	121	18.8	22.6
Continued primary education	423	19.7	23.6	53	28.7	33.8	176	20.0	24.2
No education or only primary education	183	20.0	24.2	31	28.9	34.3	63	18.9	23.5

** N = 6: missing

Age group: Kruskal-Wallis: .0001 < p < .001; Scheffé, p = .05 [a] – [b]

Marital status: Kruskal-Wallis: .001 < p < .01; Scheffé, not significant

Employment status: Kruskal-Wallis: p < .0001; Scheffé, p = .05 [a] – [b], [c] – [d], and for females only [e] - [f]

Educational level: Kruskal-Wallis: .001 < p < .01; Scheffé, not significant

Table 8.3.7

Distribution * of psychiatric caseness, diagnostic categories and groups across gender, age and urbanisation with reference to clinical judgement (with clinical judgement)

With clinical judgement	Gender			Age (yrs)					Urbanisation		
	Male N=803	Female N=996	p	18-29 N=269	30-44 N=671	45-59 N=541	60-74 N=318	p	City N=525	SC N=1274	p
Caseness	16.2	19.1	***	17.1	17.2	20.1 [a]	15.5 [b]	***	20.1	16.3	***
Psychotic disorder	0.5	0.5		1.7	0.4	0.2	0.0		1.4	0.2	
Schizophrenia	0.3	0.5		1.7	0.2	0.2	0.0		1.4	0.1	
Delusional disorder	0.2	0.0		0.0	0.2	0.0	0.0		0.0	0.1	
Mood disorder	2.2	5.1	*	2.3	3.7	4.8	3.5		6.6	2.7	**
Depression	1.3	3.9	**	1.4 [d]	2.3	4.1 [c]	2.6	*	4.7	2.0	*
Mania, hypomania, mixed disorder	0.3	0.3		0.0	0.4	0.5	0.0		0.0	0.4	
Dysthymic disorder	1.0	2.7	*	1.4	1.8	2.6	1.3		3.7	1.2	*
Anxiety disorder	3.0	6.0		4.9	4.8	4.8	3.0		5.0	4.3	
Generalised anxiety disorder	0.2	0.4		0.5	0.2	0.2	0.4		0.5	0.2	
Phobic or panic disorder	1.8	3.2		1.8	3.2	2.6	1.3		1.8	2.8	
Obsessive-compulsive disorder	0.8	1.8		2.2	1.1	1.4	0.9		1.3	1.3	
Other anxiety disorders	0.3	0.9		0.5	0.5	1.0	0.4		1.3	0.4	
Somatoform disorder	2.8	2.8		2.7	2.9	2.4	3.1		2.4	2.9	
Somatisation disorder	0.2	0.1		0.5	0.0	0.2	0.0		0.5	0.0	*
Pain disorder	1.0	1.4		0.9	1.1	0.7	2.7		1.6	1.0	
Other somatoform disorder	1.8	1.3		1.4	1.9	1.7	0.4		0.5	1.9	*
Sleep disorder	2.1	7.1	** *	1.8 [d]	4.3 [d]	6.7 [c]	5.3	**	4.5	4.8	
Dyssomnia	0.6	3.5	**	0.0	2.5	2.6	2.7		1.6	2.4	
Parasomnia	0.8	3.2	**	0.5 [d]	1.8	3.3 [c]	2.2	*	2.6	1.9	
Other sleep disorder	0.6	0.4		1.4	0.0	0.7	0.4		0.3	0.6	
Other psychiatric disorder	1.0	0.4		0.5	0.2	1.4	0.9		1.9	0.3	

* after correction for gender, age, and urbanisation

p: **caseness**: ANOVA: *: .01 < p ≤ .05; **: .001 < p ≤ .01; ***: p ≤ .001, Scheffé, p = .05) [a] – [b]

categories and groups: χ^2: *: .01 < p ≤ .05; **: .001 < p ≤ .01; ***: p ≤ .001; age: Scheffé, p = .05 [c] – [d]

Table 8.3.8 (continued)
Distribution * of psychiatric caseness, diagnostic categories and groups across gender, age and urbanisation with reference to clinical judgement (without clinical judgement)

	Gender			Age (yrs)					Urbanisation		
Without clinical judgement	Male N=803	Female N=996	p	18-29 N=269	30-44 N=671	45-59 N=541	60-74 N=318	p	City N=525	SC N=1274	p
Caseness	19.6	23.1	***	21.0	20.8	24.2 [a]	18.7 [b]	***	24.2	19.9	***
Psychotic disorder	0.6	0.8		2.2	0.5	0.2	0.4		1.9	0.3	
Schizophrenia	0.5	0.6		2.2	0.4	0.2	0.0		1.6	0.2	
Delusional disorder	0.3	0.1		0.0	0.2	0.2	0.4		0.3	0.2	
Mood disorder	4.0	6.9	*	4.1	4.9	7.6	4.3		8.6	4.3	**
Depression	3.0	5.7	**	3.2 [d]	3.3 [d]	6.9 [c]	3.9	*	6.6	3.7	*
Mania, hypomania, mixed disorder	0.3	0.4		0.0	0.5	0.5	0.0		0.3	0.4	
Dysthymic disorder	1.1	2.7	*	1.4	1.8	2.9	1.3		3.7	1.3	*
Anxiety disorder	3.1	7.5		4.9	5.5	6.6	3.5		6.6	5.0	
Generalised anxiety disorder	0.3	0.9		0.5	0.4	1.2	0.4		1.3	0.4	
Phobic or panic disorder	1.8	4.0		1.8	3.5	3.5	1.7		2.9	3.0	
Obsessive-compulsive disorder	0.8	1.8		2.2	1.1	1.4	0.9		1.3	1.3	
Other anxiety disorders	0.3	1.0		0.5	0.7	1.0	0.4		1.0	0.6	
Somatoform disorder	3.9	5.7		6.0	6.0	3.1	3.9		4.0	5.1	
Somatisation disorder	0.2	0.1		0.5	0.0	0.2	0.0		0.5	0.0	*
Pain disorder	0.8	1.3		0.5	0.9	0.7	2.7		1.0	1.0	
Other somatoform disorder	3.1	4.3		5.0	5.1	2.4	1.3		2.7	4.1	*
Sleep disorder	2.2	7.6	** *	1.4 [d]	4.5 [d]	6.9 [c]	7.1	**	4.7	5.1	
Dyssomnia	0.9	3.3	**	0.0	1.7	2.8	4.4		1.6	2.4	
Parasomnia	0.8	3.2	**	0.5 [d]	1.8	3.3 [c]	2.2	*	2.6	1.9	
Other sleep disorder	0.5	1.1		0.9	1.0	0.7	.04		0.5	0.9	
Other psychiatric disorder	1.0	1.6		0.9	1.0	2.3	0.9		2.4	1.0	

* after correction for gender, age, and urbanisation

p: **caseness**: ANOVA: *: .01 < p ≤ .05; **: .001 < p ≤ .01; ***: p ≤ .001, Scheffé, p = .05) [a] – [b]

categories and groups: χ^2: *: .01 < p ≤ .05; **: .001 < p ≤ .01; ***: p ≤ .001; age: Scheffé, p = .05 [c] – [d]

8.3.2.3 Clinical judgement and the rate of comorbidity

For the population as a whole, when clinical judgement was omitted, monomorbidity increased by 1.6%, while comorbidity increased by 2.2% (Table 8.3.9). In other words, when clinical judgement was used, a quarter (26.1%) of all subjects diagnosed as a case were comorbid, whereas this was true for one third of all subjects (31.6%) when clinical judgement was not used. No significant differences were found for the influence of clinical judgement on the distribution of the comorbidity rate.

Table 8.3.9
Distribution of diagnostic mono- and comorbidity across age, gender and urbanisation with reference to clinical judgement

		Total	Gender			Age					Urbanisation		
			male N=803	female N=996	p	18-29 N=269	30-44 N=671	45-59 N=541	60-74 N=318	p	City N=525	SC N=1274	p
With clinical judgement	Caseness	17.6	16.2	19.1	***	17.1	17.2	20.1 [a]	15.5 [b]	***	20.1	16.3	***
	1 diagnosis	13.0	12.3	13.8	**	12.7	12.8	14.4 [c]	11.9 [d]	***	14.4	12.3	***
	2 diagnoses	3.0	2.7	3.3	**	2.7	2.9	3.6 [c]	2.7 [d]	**	3.6	2.7	***
	3 diagnoses	0.7	0.6	0.8		0.5	0.7	1.0	0.6	*	1.1	0.5	***
	4 diagnoses	0.9	0.7	1.0		0.6	0.8	1.3	0.8	*	1.3	0.6	***
Without clinical judgement	Caseness	21.4	19.6	23.1	***	21.0	20.8	24.2 [a]	18.7 [b]	***	24.2	19.9	***
	1 diagnosis	14.6	13.6	15.6	**	14.2	14.3	16.4 [c]	13.1 [d]	***	16.4	13.6	***
	2 diagnoses	5.0	4.7	5.3	**	4.9	4.9	5.6 [c]	4.6 [d]	**	5.6	4.7	***
	3 diagnoses	0.9	0.7	1.0		0.6	0.8	1.3	0.7	*	1.3	0.7	***
	4 diagnoses	0.9	0.7	1.0		0.6	0.8	1.3	0.8	*	1.3	0.6	***

Caseness: ANOVA: *: .01 < p ≤ .05; **: .001 < p ≤ .01; ***: p ≤ .001, Scheffé, p = .05 [a] - [b]
Number of diagnoses per subject: χ^2: *: .01 < p ≤ .05; **: .001 < p ≤ .01; ***: p ≤ .001;
age: Scheffé, p = .05 [c] - [d]

8.4 Discussion

8.4.1 Study limitations

Our findings on the effect of the clinical judgement in the SCAN are an underestimation for two reasons. First, because we were not able to operationalise all other forms of clinical judgement (as explained in the introductory part of this chapter), this contributed to the underestimation.
Second, the SCAN is a comprehensive interview, composed of sections each of which covers a specific area of signs and symptoms. Each section starts with a brief set of core questions that screen for the probability of positive signs and symptoms in the second, extensive part of the section in question. As a consequence, when the interviewer is convinced that the remaining part of the section will not yield any major symptoms relevant to the assessment of a possible diagnosis, the rest of the section is omitted. As the primary goal of the NHA-2 project was to carry out an epidemiological study, this rule of thumb was formally followed, with the result that not every possible symptom was scored for every individual subject. Consequently, the differences between prevalence rates with and without clinical judgement probably are somewhat underestimated.

8.4.2 The difference with and without clinical judgement

Without clinical judgement, the number of psychiatric cases among the subjects we interviewed in T2 of the survey increased by 25%. The various diagnostic categories show a differentiated sensitivity to clinical judgement. In addition to an increased number of cases (pronounced for somatoform and sleep disorders), in some categories the additional diagnoses represented a qualitative shift of specific diagnoses, especially for mood and anxiety disorders. Three findings can be distinguished between the situations with and without clinical judgement: no difference; an increase in positive cases; and a change of positive cases of lower into major diagnostic categories, higher in hierarchy (see Table 8.3.1).
An explanation for the first finding, as with substance-related, eating and dissociative disorders (no difference), is that in the DSM-IV and the SCAN, for these three categories, the important symptoms constituting the diagnostic criteria are based mainly on for the subject or for the interviewer quantifiable, observable, behavioural features and not on feelings and thoughts. Because a specific behaviour is either present or absent, there is hardly any room for clinical judgement.
The second finding, as for example with mood and anxiety disorders (an increase in positive cases), is what would be expected when clinical judgement

is omitted. In these diagnostic categories, many of the diagnostic criteria are based on symptoms concerning feelings and thoughts, which make them sensitive to clinical judgement[xi]. Furthermore, these DSM criteria are often composed of a complex of possible symptoms, which also increases the influence of clinical judgement.

Another large increase when clinical judgement is left out was found for somatoform disorder. Of all the subjects with an additional somatoform disorder, 22 (81%) changed from non-morbidity to a case, whereas they only had a single common experience, now counted as a symptom. Clinical judgement clearly had a corrective effect on this perceived morbidity.

The third finding, as with somatoform and sleep disorders (an increase in positive specific diagnoses combined with a loss of these diagnoses) reflects two different mechanisms. The fact that additional positive cases were found is part of the second finding. But, for example, at the same time, of the 48 positive subjects with a sleep disorder as assessed with the formal version of the SCAN, nine lost their sleep disorder when clinical judgement was omitted. This finding can be traced to the hierarchy in the DSM-IV ("is not better accounted for by…"), which tries to locate specific symptoms as part of one, and no more than one, diagnosis. For instance, symptoms can be either part of sleep or somatoform disorders or part of other, hierarchically 'higher' diagnoses, like depression (floating symptoms). For some subjects who had only just failed to reach the threshold for depression when clinical judgement was used, a floating symptom like insomnia can be attributed to a sleep disorder and pain to a somatoform disorder. With the omission of clinical judgement, one or more subclinical symptoms pushed depression over its threshold, resulting in 25 subjects being diagnosed with a mood disorder, mainly depression, which was not assessed with the formal version of the SCAN. At the same time, this increase in depressive cases was at the expense of a loss of positive cases for the diagnostic category lower in the hierarchy.

8.4.3 The consequences of omitting clinical judgement for epidemiological data

When clinical judgement was omitted, the prevalence rate of psychiatric disorders increased significantly by 3.8% ($p \leq .001$) in comparison with the rate when the formal version of the SCAN was used. With respect to diagnostic groups, the increase was mainly attributable to somatoform disorders NOS and depression. For the other diagnostic groups, the power was insufficient to establish a significance level for the increase (for example, for schizophrenia, delusional disorder and phobic and panic disorder) or there was no difference at all (substance dependence and abuse, anorexia nervosa and dissociation disorder).

It can be concluded from this study that clinical judgement does not affect the distribution of psychiatric disorders in the general population. Nearly all increases were proportional to the prevalence rate established with the formal version of the SCAN.
With respect to comorbidity, without clinical judgement the absolute number and the percentage of the comorbid subjects increased by more than for the monomorbid subjects. In addition to the 25% increase in caseness, omitting clinical judgement had an even larger effect on the qualitative shift of an additional specific diagnosis for those subjects who were already a case.

8.4.4 Conclusions

All in all, clinical judgement interfered significantly with the prevalence rates found.
First, the prevalence rate of caseness increased by one quarter when clinical judgement was omitted from the SCAN. Second, not all specific diagnoses are equally sensitive to clinical judgement as operationalised in the SCAN. Diagnoses based on subjective experiences are more sensitive than those based on behaviour. Third, when clinical judgement was left out, the specific diagnoses of lower hierarchical ranking of those subjects who were already a case, tended to shift towards the major diagnostic categories. Fourth, some specific diagnoses (like somatoform disorder NOS) are extremely sensitive to clinical judgement, because the presence or absence of one single sign or symptom (related to daily life troubles) is crucial. Consequently, it is imperative that more specific research should be done into the meaning and the significance of clinical judgement in the diagnostic process in psychiatry (Table 8.1.1). Also, in any new research into clinical judgement in the SCAN, in order to avoid underestimation, the routine practice of omitting the second part of sections for which core questions do not yield positive symptoms should be abandoned. Finally, as clinical judgement is the gold standard to make a difference between illness and disease (see chapter 3), more research is needed to carefully unravel the way in which the answer of the subject turns into the rating of the interviewer.

References

Andrews G, Peters L, Guzman AM and Bird K. (1995) A comparison of two structured diagnostic interviews: CIDI and SCAN. *Australian and New Zealand Journal of Psychiatry,* 29, 124-132.

Beals J, Novins DK, Spicer P, Orton HD, Mitchell CM, Baron AE, Manson SM, Big Crow CK, Buchwald D, Chambers B, Christensen ML, Dillard DA, DuBray K, Espinoza PA, Flemming CM, Frederick AW, Gurley D, Jervis LL, Jim SM, Kaufman CE, Keane EM, Klein SA, Lee D, McNutly MC, Middlebrook DL, Moore LA, Nez TD, Norton IM, Randall CJ, Sam A, Shore JH, Simpson SG and Yazzie LL. (2004) Challenges in operationalizing the DSM-IV clinical significance criterion. *Archives of General Psychiatry,* 61, 1197-1207.

Brugha TS, Jenkins R, Taub N, Meltzer H and Bebbington PE. (2001) A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychological Medicine,* 31, 1001-1013.

Brugha TS, Nienhuis F, Bagchi D, Smith J and Meltzer H. (1999) The survey form of SCAN: the feasibility of using experienced lay survey interviewers to administer a semi-structured systematic clinical assessment of psychotic and non-psychotic disorders. *Psychological Medicine,* 29, 703-711.

CBS. (1997) *Census data 1997*. Voorburg/Heerlen.

Cheng AT, Tien AY, Chang CJ, Brugha TS, Cooper JE, Lee CS, Compton W, Liu CY, Yu WY and Chen HM. (2001) Cross-cultural implementation of a Chinese version of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) in Taiwan. *British Journal of Psychiatry,* 178, 567-572.

Duine TJ, Hodiamont PPG, Peer PGM and Boerma LH. (1985) Reliability aspects of the Present State Examination in four different populations [in Dutch]. *Tijdschrift voor Psychiatrie,* 27, 326-340.

Dunn G, Pickles A, Tansella M and Vazquez-Barquero JL. (1999) Two-phase epidemiological surveys in psychiatric research [editorial]. *British Journal of Psychiatry,* 174, 95-100.

Efron B and Tibshirani RJ. (1993) *Introduction to the bootstrap*. (Vols. 57) Chapman and Hall, New York.

Giel R and Nienhuis FJ. (1997) *SCAN-2.1: Schedules for clinical assessment in neuropsychiatry [in Dutch]*. WHO, Geneva / Groningen.

Henderson S, Duncan-Jones P, Byrne DG, Scott R and Adcock S. (1979) Psychiatric disorder in Canberra. A standardised study of prevalence. *Acta Psychiatrica Scandinavica,* 60, 355-374.

Hodiamont PPG and Veling SHJ. (1984) A model to assess psychiatric prevalence: The relation GHQ-PSE. [in Dutch]. *Tijdschrift voor Psychiatrie,* 26, 592-601.

Kendell RE. (1975) Psychiatric diagnosis in Britain and the United States. *British Journal of Psychiatry,* Spec No 9, 453-461.

Kendell RE, Cooper JE, Gourlay AJ, Copeland JR, Sharpe L and Gurland BJ. (1971) Diagnostic criteria of American and British psychiatrists. *Archives of General Psychiatry,* 25, 123-130.

Lesage AD, Cyr M and Toupin J. (1991) Reliable use of the Present State Examination by psychiatric nurses for clinical studies of psychotic and nonpsychotic patients. *Acta Psychiatrica Scandinavica,* 83, 121-124.

Murphy JM. (2002) Symptom scales and diagnostic schedules in adult psychiatry. In MT Tsuang en M Tohen (Eds.), *Textbook in psychiatric epidemiology, second edition*. pp. 273-332. Wiley-Liss, New York.

Narrow WE, Rae DS, Robins LN and Regier DA. (2002) Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 surveys' estimates. *Archives of General Psychiatry,* 59, 115-123.

Rijnders CATh, van den Berg JF, Hodiamont PP, Nienhuis FJ, Furer JW, Mulder J and Giel R. (2000) Psychometric properties of the schedules for clinical assessment in neuropsychiatry (SCAN-2.1). *Social Psychiatry and Psychiatric Epidemiology,* 35, 348-352.

Sartorius N, Shapiro R, Kimura M and Barrett K. (1972) WHO international pilot study of schizophrenia. *Psychological Medicine,* 2, 422-425.

Surtees PG, Dean C, Ingham JG, Kreitman NB, Miller PM and Sashidharan SP. (1983) Psychiatric disorder in women from an Edinburgh community: associations with demographic factors. *British Journal of Psychiatry,* 142, 238-246.

Tomov T and Nikolov V. (1990) Reliability of SCAN categories and scores. Results of the field trials. In CN Stefanis (Ed.), *Psychiatry: a world perspective*. pp. 107-117. Elsevier Science Publishers, Amsterdam.

Ward CH, Beck AT, Mendelson M, Mock JE and Erbauch JK. (1962) The psychiatric nomenclature. Reasons for diagnostic disagreement. *Archives of General Psychiatry,* 7, 198-205.

Wing JK. (1996) SCAN and the PSE tradition. *Social Psychiatry and Psychiatric Epidemiology,* 31, 50-54.

Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D and Sartorius N. (1990) SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry,* 47, 589-593.

Wing JK, Nixon JM, Mann SA and Leff JP. (1977) Reliability of the PSE (ninth edition) used in a population study. *Psychological Medicine,* 7, 505-516.

Wittchen HU, Üstün TB and Kessler RC. (1999) Diagnosing mental disorders in the community. A difference that matters? [editorial]. *Psychological Medicine,* 29, 1021-1027.

9 A comparison of the composite international diagnostic interview (CIDI) with the schedules for clinical assessment in neuropsychiatry (SCAN) in the general population

C.A.Th. Rijnders, MD[1,2], G.L. Van Heck[3], J.W. Furer[1] and P.P.G. Hodiamont[1,4]
[1] Dept. of Social Medicine, Radboud University Nijmegen Medical Center
[2] Institute for Mental Health Care Midden-Brabant, Tilburg
[3] Dept. of Medical Psychology, Tilburg University, Tilburg
[4] Dept. of Developmental and Clinical Psychology, Tilburg University, Tilburg

9.1 Introduction

Recent epidemiological investigations of psychiatric disorders have shown large discrepancies in prevalence rates (Andrews et al. 1995; Brugha et al. 2001; Narrow et al. 2002; Surtees et al. 1983) between studies that use the Composite International Diagnostic Interview [CIDI] as a diagnostic interview (Robins et al. 1988) and those that use the Schedules for Clinical Assessment in Neuropsychiatry [SCAN] (Wing et al. 1990).

The CIDI is a comprehensive, fully-structured diagnostic interview without clinical judgment that maps psychiatric symptoms onto DSM-IV and ICD-10 diagnostic criteria in a top-down approach (Üstün & Tien 1995). Because it is fully structured, the CIDI is suitable for application in large epidemiological studies in which well trained lay survey interviewers are used.

The SCAN-2.1 is a semi-structured, bottom-up clinical diagnostic interview developed by the World Health Organization (WHO), which includes the latest (10^{th}) version of the Present State Examination (PSE-10), the successor to the PSE-9 (Wing et al. 1990; Wing 1996). The SCAN explicitly provides scope for interviewer probes and interviewer judgement with reference to rating signs and symptoms as present or absent. The SCAN-2.1 requires clinically experienced trained interviewers and is primarily designed for use by psychiatrists and clinical psychologists. The interview is time-consuming and consequently less suitable for large-scale epidemiological surveys.

An extensive literature on the CIDI (Sartorius & Janca 1996) has established that its inter-rater reliability is good to excellent (Cottler et al. 1991; Cottler et al. 1997; Wittchen et al. 1991), its test-retest reliability good (Wittchen 1994) and its validity satisfactory to good (Peters et al. 1996).

Kurdyak and Gnam (2005), focussing on the depression module, pointed out that the CIDI has never been validated in a general population sample and interpret this fact as a major flaw in its development. Until now, validity studies have been done in clinical samples, in which the depression prevalence rate is substantially higher than the rate in community samples. Kurdyak and Gnam (2005) state that the lower prevalence of psychiatric disorders in a community-based setting most likely results in a high rate of false positives and consequently in inflated prevalence rates when the CIDI is used.

There is less literature on the SCAN than on the CIDI. Nevertheless, the claim that clinical interviews like the SCAN-2.1 are imperfect because of low inter-rater reliability and potential validity problems (Wittchen et al. 1999) has been rebutted for the PSE-9 (Duine et al. 1985b; Lesage et al. 1991; Wing et al. 1977) and the SCAN-2.1 (Andrews et al. 1995; Brugha et al. 1999b; Cheng et al. 2001; Rijnders et al. 2000; Tomov & Nikolov 1990). Globally, the reliability of the SCAN-2.1 can be described as moderate to substantial. For the test-retest procedure, reliability has been shown to be fair to moderate. In the standardised procedure agreement has been found to be substantial to almost perfect (Rijnders et al. 2000).

There is a scarcity of studies analysing diagnostic agreement between the CIDI and the SCAN. Using confirmatory SCAN interviews and the CIDI as a screening method for identifying subjects with schizophrenia and affective disorders in a community survey, Shibre et al. (2002) found that the performance of the CIDI was modest; in particular, the CIDI identified only 55% of the subjects with schizophrenia and, slightly better, only 66% of the subjects diagnosed with an affective disorder. Compton et al. (1996) reported fair to good CIDI-SCAN agreement for alcohol and cocaine dependence, but only poor agreement for opiate and cannabis dependence. Pull et al. (1997) also observed fair to good agreement between the CIDI and the SCAN for ICD-10 diagnoses of substance dependence, but extremely poor agreement for diagnoses of substance abuse. In an early study, Andrews et al. (1995; Andrews & Peters 1998) compared agreement for diagnoses of depression on the basis of independent CIDI and SCAN interviews. They found that this was moderate and that more diagnoses were attributed by the CIDI than by the SCAN. Andrews et al. (1995) expressed doubt about the appropriateness of Cohen's kappa coefficient (κ). Consequently, they suggested the use of canonical correlation analyses. Following their own advice, they obtained for the 1-month timeframe three significant, albeit rather moderate, canonical correlations – 0.69, 0.59, and 0.44, and for the lifetime timeframe also three significant canonical correlations – 0.66, 0.65, and 0.42. Basing themselves primarily on the latter findings, they concluded that the diagnostic distinctions made with the CIDI and the SCAN are highly similar.

More recently, in a sample of non-clinical young adults in Finland, Aalto-Setälä et al. (2002a) investigated diagnostic accuracy in the assessment of depression with the short form of the CIDI [CIDI-SF] (Kessler et al. 1998) and the SCAN 2.0 (WHO 1994). They found a modest correspondence between the CIDI-SF major depression episode (MDE) and the SCAN-1 (consensus) MDE: nearly one-third (28%) of the participants diagnosed with the SCAN MDE were not identified with the CIDI-SF MDE, and more than half (55%) of the participants diagnosed with MDE with the CIDI-SF were not identified with the SCAN. The κ for the correspondence of the two instruments with reference to the MDE was 0.43. In the analyses of the correspondence of the CIDI-SF MDE to the SCAN consensus, the κ value was substantially higher (0.60) for any affective disorder. These outcomes suggest that the CIDI-SF identifies a broader kind of affective disorder. Consistent with these outcomes are the results of the CIDI-SCAN comparisons reported by Brugha et al. (2001), who obtained a κ of 0.15 for depressive episode and a κ of 0.39 for any mood disorder.

In a recent study, Jordanova et al. (2004) compared the psychometric properties of the CIDI with those of the SCAN, conventionally considered the gold standard, in a sample of primary care recipients. The agreement for depressive episodes was moderate: $\kappa = 0.54$. In line with the results of the study by Andrews et al. (1995), the CIDI showed a marked tendency to over-diagnose, resulting in a prevalence of 18.1%, whereas the corresponding figure with the SCAN was 7.6%. Agreement was excellent for any depression, as established by a κ value of 0.97.

In conclusion, the available evidence is far from decisive. Consequently, we agree fully with Jordanova et al. (2004), who concluded that much more research is required to determine the extent and limits of the validity of the CIDI. We should add that the last word has not been said regarding the psychometric and clinimetric qualities of the SCAN and the CIDI-SCAN correspondences.

This study was undertaken in order to improve our understanding of the common finding that lay and clinical psychiatric diagnostic interviews administered in the general population consistently furnish discrepant findings, which suggests that they yield different information. We focused our study on depressive disorders for two reasons: first, these disorders form one of the diagnostic groups that are very sensitive to clinical judgement, probably because diagnosis criteria are based on subjective experiences rather than on objective behaviour; and second, in the data that we gathered for this CIDI-SCAN comparison, more depressive disorders were found than any other diagnostic group, which provided the basis for a meaningful comparison.

9.2 Methods

9.2.1 Setting

This study was conducted as part of the Nijmegen Health Area 2 (NHA-2) project, a general population survey that was carried out in this area (NHA, the Netherlands) with the Dutch version of the SCAN-2.1 in 1997.

9.2.2 Instruments and design

In this study, we used the CIDI-2.1 auto version of August 1997 and the preliminary computerised SCAN-2.1 interview version of July 1997. Consistent with the clinical, semi-structured character of the SCAN, 30 psychologists and psychiatric nurses were selected for their knowledge of psychopathology. Special sessions on psychiatric phenomenology were organised to optimise their clinical judgement. All interviewers participated in an intensive 1-week training course given by the WHO-SCAN collaborating centre in Groningen so that they could learn to work with the pencil-and-paper version of the SCAN. A total of 10 booster sessions were held to improve their interviewing skills. Furthermore, a WHO-approved computerised version of the SCAN-2.1 was introduced, inclusive of an algorithm for determining DSM-IV diagnoses. During the fieldwork, problems that arose while the data were being checked were discussed with a psychiatrist. The CIDI was administered by 13 lay interviewers who had also participated in the 1996 wave of the NEMESIS study and were trained by the WHO-CIDI centre in Amsterdam. To avoid any differences due to the sequence of the two interviews, the study was carried out in counterbalanced order. In a first wave in the NHA, a sample of 70 subjects stratified by age, gender, extent of urbanisation and GHQ-score who had been interviewed as part of the NHA-2 sample with the SCAN-2.1 were interviewed with the CIDI-2.1 within a week after the initial interview. In a second wave a sample of 69 subjects were selected in the Tilburg Health Area in accordance with the same stratification procedure as that used for the Nijmegen wave. These participants were interviewed in the reverse order (first the CIDI, then the SCAN, within a 1-week timeframe). The two health areas are comparable in that they both include one larger city with a large student population and a number of surrounding communities that are urbanised to varying degrees. None of the interviewers conducted both interviews with the same patient.

9.2.3 Participants

The subsample of 70 subjects who were evaluated first with the SCAN and then the CIDI contained 28 males and 42 females. The average age of the males was 50.4 years (range 26 - 75; SD = 13.0) and of the females 47.6 years (range 21 - 73; SD = 15.0). The subsample of 69 participants who were evaluated first

with the CIDI and then with the SCAN contained 27 males and 42 females. The average age of the males was 48.4 years (range 20 - 75; SD = 16.4) and of the females 52.4 years (range 19 - 75; SD = 14.8).

9.2.4 Statistics

Extent of agreement and patterns of agreement were studied with kappa (κ) coefficients (Cohen 1968) and canonical correlation analysis, respectively. Following Andrews et al. (1995), we conducted our analyses of the degree of agreement between the CIDI and the SCAN mainly with κ coefficients. The reason for doing these analyses is that at this stage of development of the two diagnostic methods we are not yet in a position to conclude which of the two assessment instruments can be considered the standard against which the other should be evaluated. The relationship between the two sets of variables was further examined by means of canonical correlation analysis [see Tabachnik & Fidell (1996) for an overview]. Canonical correlations establish the degree of association between two latent variables (canonical variates). In this study, one variate represents a set of CIDI items and the other a set of SCAN items. There may be more than one significant canonical correlation and each canonical correlation is optimised so that the linear correlation between the two latent variables is maximised. The process is repeated until a successive linear combination is no longer significant.

In the canonical correlation program of SPSS version 13.0 (CANCORR) that was used, the size of each set of items is restricted to a maximum of 25. As a consequence, we had to reduce the number of items by aggregating those that target loss of energy and sleep problems (SCAN), and thinking disturbances, suicidal ideation, self-confidence and appetite changes (CIDI); only the items of the basic diagnosis of the depressive episode according to the DSM-IV were used. The explicit observational items of the SCAN (n = 15) were excluded because there was no counterpart to these in the CIDI.

9.3 Results

The CIDI identified 39 subjects with a psychiatric disorder (case), while the SCAN identified 34 such subjects. Of the 39 subjects with a CIDI diagnosis, the SCAN identified 22, leaving 17 subjects undiagnosed. Of the 100 subjects with no CIDI diagnosis, 12 were indicated by the SCAN as a case. It follows that 88 subjects were not diagnosed by either method. At 0.46, the κ value was modest. Re-scoring of the SCAN by ignoring the clinical judgements raised the number of cases to 47. In that case, 29 subjects diagnosed with the CIDI were also diagnosed by the SCAN, so that 10 subjects were undiagnosed. Of the 100 subjects without caseness according to the CIDI, the SCAN identified 22, i.e. 78 subjects were not identified by either method. The percentage of SCAN cases that were picked up by the CIDI decreased from 64.7% to 61.7%, while the percentage of CIDI cases that were also diagnosed using the SCAN increased from 56.4% to 74.4%. In spite of this, at 0.53, the κ value was modest.

With reference to the diagnostic categories, agreement of the CIDI with the original SCAN was in the range 0.09 – 0.51, the lowest κ having been found for somatisation disorders and the highest for mood disorders. κ coefficients for disorders related to substance use and anxiety disorder were 0.27 and 0.35, respectively. When the clinical judgement of the SCAN was omitted, there was only a modest increase in κ for mood disorders (κ = 0.54).

Agreement analyses of the items involving comparisons of corresponding CIDI items and original SCAN items yielded κ values ranging from -0.03 to 0.53 (see Table 9.3.1). These coefficients were in the poor to fair range (Landis & Koch 1977b)[xii]. When the clinical judgement of the SCAN was left out, the comparisons of corresponding CIDI and SCAN items yielded κ values ranging from 0.14 to 0.41 (see Table 9.3.2). Removing the clinical judgment somewhat raised the correlations; all κ coefficients reached significance, while the average κ value reached 0.28, in comparison with a mean κ of 0.27 for the original scores on the SCAN.

Table 9.3.1
Agreement on the diagnosis of individual depression symptoms between the CIDI and the original SCAN

CIDI symptoms	SCAN symptoms	Kappa	95%-CI	p ≤
Moving all the time (E11)	Subjective restlessness (3.006)	0.31	± 0.24	0.001
Lack of energy/feeling tired (E3)	Fatigability and exhaustion; loss of energy; feelings of being overwhelmed by daily tasks (3.007, 7.006, 7.007)	0.36	± 0.18	0.001
Periods of feeling sad, empty or depressed (E1)	Depressive mood (6.001)	0.53	± 0.18	0.001
Loss of interest in most matters (E2)	Anhedonia (6.004)	0.41	± 0.20	0.001
Frequent thoughts about death (E18)	Preoccupation with death/disaster (6.010)	0.20	± 0.27	0.005
Frequent thoughts about committing suicide (E19)	Suicide or self-inflicted damage (6.011)	0.39	± 0.55	0.001
Guilt feelings (E12A)	Pathological guilt (6.013)	0.27	± 0.27	0.001
Inferiority feelings, low self-confidence (E13, E14)	Loss of self-confidence (6.015)	0.19	± 0.20	0.009
Feelings of worthlessness (E12)	Loss of self-respect (6.017)	0.24	± 0.21	0.001
Trouble concentrating (E15)	Loss of concentration (7.002)	0.25	± 0.21	0.003
Trouble thinking: slow thinking, indecisiveness (E16,E17)	Subjective, inefficient thinking (7.003)	0.19	± 0.19	0.006
Inability to enjoy good fortune (E24)	Loss of interests (7.004)	0.25	± 0.22	0.002
Slowed talking or movements (E10)	Subjective, experienced slowness (7.005)	0.13	± 0.23	0.002
Appetite changes (E4,E6)	Appetite changes (8.005)	0.22	± 0.23	0.007
Weight loss (E5)	Body weight loss (8.006)	-0.03	± 0.02	*ns*
Weight gain (E7)	Body weight gain (8.007)	0.32	± 0.49	0.001
Trouble sleeping (E8)	Difficulty getting to sleep, insomnia, early awakening (8.011, 8.013, 8.014)	0.35	± 0.19	0.001
Excessive sleeping (E9)	Hypersomnia (8.016)	0.12	± 0.27	*ns*

Table 9.3.2
Agreement on the diagnosis of individual depression symptoms between the CIDI and the SCAN without clinical judgment

CIDI symptoms	SCAN symptoms	Kappa	95%-CI	$p \leq$
Moving all the time (E11)	Subjective restlessness (3.006)	0.24	± 0.16	0.001
Lack of energy/ feeling tired (E3)	Fatigability and exhaustion; Loss of energy; feelings of being overwhelmed by daily tasks (3.007, 7.006, 7.007)	0.32	± 0.15	0.001
Periods of feeling sad, empty or depressed (E1)	Depressive mood (6.001)	0.41	± 0.13	0.001
Loss of interest in most matters (E2)	Anhedonia (6.004)	0.36	± 0.18	0.001
Frequent thoughts about death (E18)	Preoccupation with death/disaster (6.010)	0.24	± 0.24	0.005
Frequent thoughts about committing suicide (E19)	Suicide or self-inflicted damage (6.011)	0.39	± 0.55	0.001
Guilt feelings (E12A)	Pathological guilt (6.013)	0.30	± 0.25	0.001
Inferiority feelings, low self-confidence (E13, E14)	Loss of self-confidence (6.015)	0.19	± 0.20	0.009
Feelings of worthlessness (E12)	Loss of self-respect (6.017)	0.17	± 0.20	0.042
Trouble concentrating (E15)	Loss of concentration (7.002)	0.27	± 0.16	0.001
Trouble thinking: slow thinking, indecisiveness (E16,E17)	Subjective, inefficient thinking (7.003)	0.32	± 0.20	0.001
Inability to enjoy good fortune (E24)	Loss of interests (7.004)	0.28	± 0.19	0.001
Slowed talking or movements (E10)	Subjective, experienced slowness (7.005)	0.20	± 0.25	0.020
Appetite changes (E4,E6)	Appetite changes (8.005)	0.22	± 0.24	0.007
Weight loss (E5)	Body weight loss (8.006)	0.17	± 0.22	0.004
Weight gain (E7)	Body weight gain (8.007)	0.14	± 0.19	0.009
Trouble sleeping (E8)	Difficulty getting to sleep, insomnia, early awakening (8.011, 8.013, 8.014)	0.29	± 0.15	0.001
Excessive sleeping (E9)	Hypersomnia (8.016)	0.20	± 0.25	0.016

Canonical correlation analyses were performed between the set of depression-related CIDI items and the set of depression-related SCAN items with the original scoring applying the SPSS CANCORR procedure. The CIDI set included 19 items reflecting general manifestations of depression; these targeted periods in which the subject felt sad or empty or depressed and had lost interest in most matters, and also specific markers of depression, in particular lack of energy, appetite change, loss of libido, sleep disturbances, slowness and restlessness, feelings of worthlessness, guilt feelings, lack of confidence, thinking disturbances and thoughts of death. The SCAN set (19 items) targeted subjective restlessness, fatigue and exhaustion, depressive mood, anhedonia, preoccupation with death or disaster, suicide or self-inflicted

damage, pathological guilt, ideas of reference accompanied by guilt feelings, loss of self-respect and self-confidence, loss of concentration, subjective inefficient thinking, loss of interest, subjectively experienced slowness, loss of energy, feelings of being overwhelmed by daily tasks, appetite change, loss of libido, changes in body weight and sleep disturbances.

Seven canonical correlations were found that reached statistical significance and were meaningful and interpretable (see Table 9.3.3). The first canonical correlation was 0.85. With reference to the CIDI, the first pair of canonical variates had significant loadings (> 0.30) on nearly two-thirds of all CIDI items, the highest loadings (> 0.60) having occurred for the items slowed talk or slowed movement. Hence the first canonical variate in the CIDI represented the core symptoms of depressive disorder, i.e. periods of feeling sad or empty or depressed, loss of interest in most matters, lack of energy, general slowing down and appetite change. As for the SCAN, the first canonical variate indicated a variety of manifestations of depression: one quarter of the SCAN items had high loadings, the highest loading having occurred for depressive mood. The second canonical correlation (r = 0.84) established that the items feelings of worthlessness, low self-confidence, preoccupation with death and loss of pleasure had substantial loadings (> 0.50) with reference to the CIDI. These findings corresponded to substantial loadings on the SCAN items preoccupation with death and disaster and loss of self-respect. The third canonical correlation was 0.81. This pair of canonical variates showed substantial loadings for the CIDI items restless movements and guilt feelings, with an equal loading for subjective experience of slowness for the SCAN. For the fourth canonical correlation for the CIDI, the only substantial marker was related to suicide, while for the SCAN, in addition to a high loading on the suicide item, ideas of reference, pathological guilt and hypersomnia had positive loadings, whereas the items low self-confidence, loss of self-respect and appetite changes had a negative loading. The fifth pair of canonical variates represented a broad pattern of depression indicators on the CIDI, and contained five items that also had high loadings on the first canonical variate, supplemented by almost all items that did not have sufficient loadings on the first variate. The major markers of the first variate, i.e. periods of feeling sad or empty, or depressed and preoccupation with death did not contribute in this respect. As for the SCAN, the canonical variate contained both major markers of the depressive disorder, depressed mood and anhedonia, and loss of interest and subjective inefficient thinking.

Table 9.3.3

Canonical variates, canonical loadings, canonical correlation, percentual variance, and redundancies between CIDI and SCAN variables and their corresponding canonical variates

Item Number CIDI	Item	Canonical variates and canonical loadings I	II	III	IV	V	VI	VII
E1	Periods of feeling sad, empty or depressed	0.43	0.41	0.32				
E2	Loss of interest in most matters	0.35		0.37		0.49		
E3	Lack of energy or feeling tired	0.35	0.40	0.36		0.36		
E4,E6	Appetite change	0.46	0.39					
E5	Weight loss					0.40		
E7	Weight gain	0.39			-0.32			0.43
E8	Trouble sleeping		0.38	0.44		0.49		
E9	Excessive sleeping	0.39					0.56	
E10	Slowed talking or movements	0.62				0.40		
E11	Moving all the time			0.56			0.35	
E12	Feelings of worthlessness		0.59			0.41		0.30
E12A	Guilt feelings	0.37		0.51				
E13, E14	Inferiority feelings; loss of self-confidence		0.51			0.31	0.30	
E15	Trouble concentrating		0.31	0.39		0.38	0.31	
E15A	Inability to read, watch television	0.43	0.35			0.41		0.31
E16,E17	Trouble thinking	0.33	0.36					
E18	Frequent thoughts about death	0.30	0.58					
E19; E20	Feeling low, frequent thoughts about committing suicide, suicide attempt	0.36			0.53	0.38		
E24	Loss of pleasure		0.53	0.41		0.36		
Percentual variance		0.13	0.13	0.11	0.03	0.11	0.07	0.04
Redundancy		0.09	0.09	0.07	0.02	0.06	0.03	0.02

With reference to the CIDI, the sixth canonical correlation had the highest loading for the item excessive sleeping, while regarding the SCAN, the item appetite changes had the highest loading. Both instruments contained the item related with problems with self-confidence. Finally, the seventh canonical correlation was still 0.62, with, for the CIDI, a sufficient loading for weight gain, feelings of worthlessness and inability to read. With reference to the SCAN, loss of energy and weight gain had the highest positive loadings and the suicide item a negative loading.

Table 9.3.3 (continued)

Canonical variates, canonical loadings, canonical correlation, percentual variance, and redundancies between CIDI and SCAN variables and their corresponding canonical variates

Item Number	Item	Canonical variates and canonical loadings I	II	III	IV	V	VI	VII
SCAN								
3.006	Subjective restlessness			0.46				
3.007/7.006/ 7.007	Fatigability, loss of energy, feelings of being overwhelmed by daily tasks	0.36				0.41		0.36
6.001	Depressive mood	0.51	0.31	0.45		0.41		
6.004	Anhedonia			0.45		0.54	0.43	
6.010	Preoccupation with death or disaster		0.55					
6.011	Suicide or self-inflicted damage				0.31			-0.36
6.013	Pathological guilt	0.33	0.44		0.36			
6.014	Ideas of reference with guilt content				0.54			
6.015	Loss of self-confidence				-0.48		0.43	
6.017	Loss of self-respect		0.60		-0.57			
7.002	Loss of concentration			0.48				
7.003	Subjective, inefficient thinking					0.33		
7.004	Loss of interests		0.42			0.40		
7.005	Subjective experience of slowness	0.35	-0.37	0.53				
8.005	Appetite changes				-0.41		0.59	
8.006	Weight loss						0.31	
8.007	Weight gain							0.46
08.011/13/14	Sleep problems			0.32				
08.016	Hypersomnia				0.30			
Percentual variance		0.05	0.09	0.07	0.08	0.07	0.06	0.04
Redundancy		0.04	0.06	0.05	0.05	0.03	0.03	0.02
Canonical correlation coefficient		0.85	0.84	0.81	0.77	0.72	0.65	0.62

9.4 Discussion

The major objective of this study was to compare the CIDI with the SCAN in terms of case identification with respect to mental disorders in general and depressive disorders in particular. The present state of affairs reflects a remarkable heterogeneity of findings due to characteristics of the populations under study but also features of the assessment methods used (Fontenelle et al. 2006). Accordingly, we felt that the current CIDI-SCAN comparison could further epidemiological knowledge since this study was not limited to subjects in treatment, which, unfortunately, is often the case. Furthermore, this comparison enables a confrontation between two instruments that differ clearly with respect to the scope for clinical judgement, which is encouraged in the SCAN and kept to a minimum in the CIDI.

The main findings can be summarized as follows. First, the data indicated that overall there was only moderate agreement on rates of diagnosis between the two methods. Agreement regarding caseness in terms of any disorder was somewhat higher after the two methods were brought closer together through eliminating the influence of clinical judgement in the SCAN. But even after this methodological intervention, agreement was only fair. Second, comparisons of the individual diagnostic criteria showed significantly lower agreement for some criteria. Notable is the fair but relatively low agreement for somatisation disorders and the substantially better agreement for mood disorders. However, even for the latter category, agreement was moderate. Third, the focus on depression did not change this picture. Generally, κ values for individual depression criteria were poor to fair. In spite of a certain degree of face validity (e.g., slowed talk or movement [CIDI] versus subjective experienced slowness [SCAN]), κ values indicated only low agreement. Actually, moderate agreement was obtained only for the comparisons between (1) the CIDI item periods of feeling sad or empty or depressed and the depressive mood item of the SCAN and (2) the CIDI item loss of interest in most matters and the SCAN item anhedonia. Filtering the influence of clinical judgement out of the SCAN diagnoses had no effect whatsoever.

In contrast to this rather disappointing level of concordance, the CIDI/SCAN canonical correlation analysis, which was conducted to explain the relation between the two sets of latent variables, established convincingly that the CIDI and the SCAN agree substantially.

Andrews et al. (1995) suggested that there are at least two possible reasons for a lack of agreement in the assignment of individuals to particular diagnoses with the two instruments: first, they point to the possible unreliability of one or both of them; and second, they suggest that the instruments may differ in the definition of the symptoms that contribute to diagnostic categories. They conducted a rather global canonical correlation analysis based on broad diagnoses (e.g.,

agoraphobia/panic, social phobia, obsessive-compulsive disorder, depression) and concluded that the outcomes “strongly suggest that the less than perfect concordance between the CIDI and the SCAN is unlikely to be due to differing definitions of diagnoses” (p. 129). The present, much more fine-tuned canonical correlation analysis based on individual items for depression makes it possible to put their conclusion in perspective. Although we found high coefficients between latent variates for individual depression items, we were not able to confirm the strong agreement for items with the κ coefficient. It proved difficult to identify the corresponding items in the SCAN and the CIDI that represent the operationalisation of the DSM-IV diagnostic criteria for mood disorders. This divergence with reference to item operationalisation creates a fundamental mismatch between the two instruments, a mismatch that is amplified by the differences in the procedures for obtaining diagnoses.

With reference to the preceding, we refer to Spitzer and Williams (1985), who distinguish three sources of data divergence: information variance, observation and interpretation variance, and criterion variance. The first source results from the fact that diagnoses can be based upon discrepancies in the information given by patients; the second arises from differences in the ways observers collect and interpret data; and the third reflects the employment of different criteria or differences with respect to conceptualisation and operationalisation. Because standard classification systems were used, Spitzer and Williams did not consider criterion variance to be a major source of variance in CIDI-SCAN comparisons. That might apply to theoretical diagnostic criteria, but the fact that the operationalisation of these criteria diverges between the two instruments introduces quite an important source of variance. A comparison of the CIDI and the SCAN with respect to items reveals that the questions do not have identical or even highly similar wordings. Although the two instruments are meant to tap the same phenomena, the phrasing is quite different. Presumably this difference reflects the fact that the fully-structured CIDI with its relatively simplified assessment task is suitable for use by lay survey interviewers, while the more sophisticated semi-structured SCAN requires clinically experienced trained interviewers. In sum, these different operationalisations imply a source of information variance.

As for the observation and interpretation variance, unlike the CIDI, the SCAN explicitly uses clinical judgement as a criterion for the clinical relevance of the assessed phenomena, thereby knowingly introducing a considerable source of variance.

Although all three sources of variance play a role in reducing the agreement in the assignment of individuals to particular diagnoses, when the CIDI and the SCAN are used, we firmly believe that the problem of criterion variance is not yet solved. The observation and interpretation variance and the information

variance from the SCAN represent the essence of the clinical approach and so cannot be easily dismissed. As far as the CIDI is concerned, observation and interpretation are abandoned in favour of face validation of the (illness) response of the subject, which introduces information variance from the act of recording a perceived morbidity.

In conclusion, future research should be aimed at scrutinising the precise influence of those sources of variance that can be dealt with, especially the operationalisation of the theoretical diagnostic criteria of the current classification systems. Such research will pave the way for improvements in the use of both the CIDI and the SCAN.

References

Aalto-Setala T, Haarasilta L, Marttunen M, Tuulio-Henriksson A, Poikolainen K, Aro H and Lonnqvist J. (2002) Major depressive episode among young adults: CIDI-SF versus SCAN consensus diagnoses. *Psychological Medicine,* 32, 1309-1314.

Andrews G and Peters L. (1998) The psychometric properties of the Composite International Diagnostic Interview. *Social Psychiatry and Psychiatric Epidemiology,* 33, 80-88.

Andrews G, Peters L, Guzman AM and Bird K. (1995) A comparison of two structured diagnostic interviews: CIDI and SCAN. *Australian and New Zealand Journal of Psychiatry,* 29, 124-132.

Brugha TS, Jenkins R, Taub N, Meltzer H and Bebbington PE. (2001) A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychological Medicine,* 31, 1001-1013.

Brugha TS, Nienhuis F, Bagchi D, Smith J and Meltzer H. (1999) The survey form of SCAN: the feasibility of using experienced lay survey interviewers to administer a semi-structured systematic clinical assessment of psychotic and non-psychotic disorders. *Psychological Medicine,* 29, 703-711.

Cheng AT, Tien AY, Chang CJ, Brugha TS, Cooper JE, Lee CS, Compton W, Liu CY, Yu WY and Chen HM. (2001) Cross-cultural implementation of a Chinese version of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) in Taiwan. *British Journal of Psychiatry,* 178, 567-572.

Cohen J. (1968) Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin,* 70, 213-220.

Compton WM, Cottler LB, Dorsey KB, Spitznagel EL and Mager DE. (1996) Comparing assessments of DSM-IV substance dependence disorders using CIDI-SAM and SCAN. *Drug and Alcohol Dependence,* 41, 179-187.

Cottler LB, Grant BF, Blaine J, Mavreas V, Pull C, Hasin D, Compton WM, Rubio-Stipec M and Mager D. (1997) Concordance of DSM-IV alcohol and drug use disorder criteria and diagnoses as measured by AUDADIS-ADR, CIDI and SCAN. *Drug and Alcohol Dependence,* 47, 195-205.

Cottler LB, Robins LN, Grant BF, Blaine J, Towle LH, Wittchen HU and Sartorius N. (1991) The CIDI-core substance abuse and dependence questions: cross-cultural and nosological issues. The WHO/ADAMHA Field Trial. *British Journal of Psychiatry,* 159, 653-658.

Duine TJ, Hodiamont PPG, Peer PGM and Boerma LH. (1985) Reliability aspects of the Present State Examination in four different populations [in Dutch]. *Tijdschrift voor Psychiatrie,* 27, 326-340.

Fontenelle LF, Mendlowicz MV and Versiani M. (2006) The descriptive epidemiology of obsessive-compulsive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry,* 30, 327-337.

Jordanova V, Wickramesinghe C, Gerada C and Prince M. (2004) Validation of two survey diagnostic interviews among primary care attendees: a comparison of CIS-R and CIDI with SCAN ICD-10 diagnostic categories. *Psychological Medicine,* 34, 1013-1024.

Kessler RC, Andrews G, Mroczak D, Üstün B and Wittchen HU. (1998) The World Health Organization Composite International Diagnostic Interview Short Form (CIDI-SF). *International journal of methods in psychiatric research,* 7, 171-185.

Kurdyak PA and Gnam WH. (2005) Small signal, big noise: performance of the CIDI depression module. *Canadian Journal of Psychiatry,* 50, 851-856.

Landis JR and Koch GG. (1977) The measurement of observer agreement for categorical data. *Biometrics,* 33, 159-174.

Lesage AD, Cyr M and Toupin J. (1991) Reliable use of the Present State Examination by psychiatric nurses for clinical studies of psychotic and nonpsychotic patients. *Acta Psychiatrica Scandinavica,* 83, 121-124.

Narrow WE, Rae DS, Robins LN and Regier DA. (2002) Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 surveys' estimates. *Archives of General Psychiatry,* 59, 115-123.

Peters L, Andrews G, Cottler LB, Chatterji S, Janca A and Smeets RMW. (1996) The Composite International Diagnostic Interview Post Traumatic Stress Disorder module: Preliminary data. *International journal of methods in psychiatric research,* 6, 167-174.

Pull CB, Saunders JB, Mavreas V, Cottler LB, Grant BF, Hasin DS, Blaine J, Mager D and Üstün BT. (1997) Concordance between ICD-10 alcohol and drug use disorder criteria and diagnoses as measured by the AUDADIS-ADR, CIDI and SCAN: results of a cross-national study. *Drug and Alcohol Dependence,* 47, 207-216.

Rijnders CATh, van den Berg JF, Hodiamont PP, Nienhuis FJ, Furer JW, Mulder J and Giel R. (2000) Psychometric properties of the schedules for clinical assessment in neuropsychiatry (SCAN-2.1). *Social Psychiatry and Psychiatric Epidemiology,* 35, 348-352.

Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA and others. (1988) The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry,* 45, 1069-1077.

Sartorius N and Janca A. (1996) Psychiatric assessment instruments developed by the World Health Organization. *Social Psychiatry and Psychiatric Epidemiology,* 31, 55-69.

Shibre T, Kebede D, Alem A, Negash A, Kibreab S, Fekadu A, Fekadu D, Jacobsson L and Kullgren G. (2002) An evaluation of two screening methods to identify cases with schizophrenia and affective disorders in a community survey in rural Ethiopia. *International Journal of Social Psychiatry,* 48, 200-208.

Spitzer RL and Williams JBW. (1985) Classification of mental disorders. In HI Kaplan en BJ Sadock (Eds.), *Comprehensive textbook of psychiatry IV*. pp. 591-613. Willams and Wilkins, Baltimore.

Surtees PG, Dean C, Ingham JG, Kreitman NB, Miller PM and Sashidharan SP. (1983) Psychiatric disorder in women from an Edinburgh community: associations with demographic factors. *British Journal of Psychiatry,* 142, 238-246.

Tabachnik BG and Fidell LS. (1996) *Using multivariate statistics*. (3rd ed.) HarperCollins College Publishers, New York.

Tomov T and Nikolov V. (1990) Reliability of SCAN categories and scores. Results of the field trials. In CN Stefanis (Ed.), *Psychiatry: a world perspective*. pp. 107-117. Elsevier Science Publishers, Amsterdam.

Üstün TB and Tien AY. (1995) Recent developments for diagnostic measures in psychiatry. *Epidemiologic Reviews,* 17, 210-220.

WHO. (1994) *SCAN: Schedules for Clinical Assessment in Neuropsychiatry, Version 2.0.* Psychiatric Publishers International/American Psychiatric Press Inc., Geneva.

Wing JK. (1996) SCAN and the PSE tradition. *Social Psychiatry and Psychiatric Epidemiology,* 31, 50-54.

Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D and Sartorius N. (1990) SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry,* 47, 589-593.

Wing JK, Nixon JM, Mann SA and Leff JP. (1977) Reliability of the PSE (ninth edition) used in a population study. *Psychological Medicine,* 7, 505-516.

Wittchen HU. (1994) Reliability and validity studies of the WHO--Composite International Diagnostic Interview (CIDI): a critical review. *Journal of Psychiatric Research,* 28, 57-84.

Wittchen HU, Robins LN, Cottler LB, Sartorius N, Burke JD and Regier D. (1991) Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI). The Multicentre WHO/ADAMHA Field Trials [published erratum appears in Br J Psychiatry 1992 Jan;160:136]. *British Journal of Psychiatry,* 159:645-53, 658, 645-53, 658.

Wittchen HU, Üstün TB and Kessler RC. (1999) Diagnosing mental disorders in the community. A difference that matters? [editorial]. *Psychological Medicine,* 29, 1021-1027.

Part IV

10 Research results and general discussion

In this chapter the empirical data (parts II and III) will be viewed from the perspective of the theoretical principles of psychiatric epidemiology and diagnostics (part I). At the end of this chapter, recommendations are given on how to improve various aspects of the diagnostic process in psychiatric epidemiology.

10.1 Case definition, case identification and case finding in psychiatric epidemiology[xiii]

Epidemiology emerged from the need to bring under control the large epidemics of the second half of the 19th century. This new discipline was very successful in that respect since identifying cases was relatively easy as far as infectious diseases were concerned. Matters got complicated, however, when epidemiological principles were applied to non-infectious and chronic diseases, particularly in psychiatry (Fleming & Hsieh 2002), because the definition of a case was not that clear.

In psychiatric epidemiology, the solutions to the problems of defining and identifying cases came in three generations (Dohrenwend & Dohrenwend 1982). The first generation of psychiatric epidemiological studies tried to avoid the complexity of the psychiatric case definition by identifying cases by means of data on hospitalised patients obtained from non-standardised, non-structured[xiv], so-called free interviews. In the absence of a clear psychiatric classification system, the second generation used questionnaires to measure general distress as a synonym for psychiatric caseness in general population surveys. The third generation came in two waves: in a first group of studies, semi-structured, standardised diagnostic clinical interviews (like the Present State Examination, ninth version [PSE-9]) were conducted in the general population, but the data from these studies had to be classified in accordance with non-criterion-based definitions of the classification systems, such as the IDC-9[xv] and the DSM-II[xvi]. The data generated by the second wave of the third generation of psychiatric epidemiological studies were classifiable on the basis of criterion-based definitions of the DSM-III/DSM-IV and the ICD-10 classification systems. Besides improving the semi-structured interviews, these classification systems cleared the way for the development of fully structured, non-clinical, standardised diagnostic interviews, like the Diagnostic Interview Schedule (DIS).

The oldest structured diagnostic interview tradition, which was in vogue in the first wave of the third generation of psychiatric epidemiological studies, was

represented by the Present State Examination/Schedules for Clinical Assessment in Neuropsychiatry (PSE/SCAN), which was first published in 1974 for use in epidemiological surveys (Sartorius et al. 1972; Wing et al. 1967; Wing et al. 1974; Wing 1996). The main characteristic of this tradition is the clinical perspective of the case identification, which is in accordance with the requirement of Cooper (1979), who stated that it is the psychiatric epidemiologist who should choose (diagnostic) variables that are not only scientifically, but also clinically relevant, and that represent a need – as opposed to a demand – for care. The second interview tradition started in response to the limited usability (low cost effectiveness) of the clinical instruments in large-scale general population surveys. The development of the DIS began in 1978 at the request of the National Institute of Mental Health (NIMH). This body needed a low-cost, comprehensive diagnostic instrument for use by lay interviewers for its large-scale, multi-centre Epidemiological Catchment Area (ECA) study. Consequently, they required a fully structured and non-clinical instrument with comprehensive specifications of the classification criteria in the questions (Eaton et al. 1981). Even though the criterion-based classification systems used in the diagnostic interviews of the two traditions are the same, they differ considerably in the prevalence rates they yield. This fact raises questions about the quality of case finding and argues for greater interchangeability of the set of instruments used in psychiatric epidemiology.

In order to get reliable and valid results, the ‘what’ in case identification (the diagnostic process) should be clear and unambiguously defined. Kendell (1993) recognised the importance of three methodological principles for improving the reliability of the diagnostic process:

1. Structured and standardised instruments should be used;
2. Any algorithms that are used should be transparent and unambiguous;
3. All signs and symptoms that are determined should be defined in such a way that the rater understands the definitions.

In the process of the development of the two diagnostic interview traditions (PSE/SCAN and DIS/CIDI), these three principles have been implemented in these traditions, but with a different timing and to a different extent. Both the DIS/CIDI and the PSE/SCAN tradition are characterised by structured and standardised instruments. With regard to the transparency and non-ambiguity of the algorithm used, the PSE/SCAN tradition has undergone appreciable improvement since the ‘black box’ of the PSE-9, whereas the transparency of the algorithm of the DIS/CIDI tradition is not yet up to scratch. With regard to the unambiguous definition of signs and symptoms, the PSE/SCAN, with the SCAN-2.1, clearly has an advantage over the DIS/CIDI tradition. The SCAN comes with an extensive and well defined glossary, while in the CIDI interview

the definition of the signs and symptoms depends on the respondent's comprehension of the terms used in the formulation of the compulsory questions.

The purpose of this study was to contribute to these areas of interest through presenting the epidemiological 1997 case-finding data of the Nijmegen Health Area project-2 (NHA-2) and comparing these data with the 1983 data of NHA-1. With respect to case identification, in this study we explored the effect of clinical judgement and compared data from the SCAN with data from the Composite International Diagnostic Interview (CIDI), the successor to the DIS[xvii]. We also elaborated on the psychometric properties of the SCAN and the differences between the SCAN and the CIDI.
The decision to use the SCAN-2.1 as the diagnostic instrument of choice in the NHA-2 made it possible to attain the following objectives:

- compare the NHA-2 data with the data of the NHA-1 of 1983 by converting PSE-10 scores into PSE-9 scores (case finding [ICD-9], monitoring [1983-1997], see 10.2.1);
- study the consequences of the relative extension of the diagnostic range by determining diagnoses on the basis of the integral diagnostic axis 1 range of the DSM-IV (case finding [DSM-IV], mapping [1997], see 10.2.2);
- study the reliability and the validity of the application of the semi-structured, clinically oriented SCAN-2.1 interview in open population surveys (case identification, see 10.3.1);
- investigate the influence of clinical judgement in the diagnostic process by using the item scores along with their qualitative clinical perspective (case identification, see 10.3.2);
- study the agreement and the differences between the results of the SCAN and the CIDI when these are obtained from the same survey sample of the open population (case identification, see 10.3.3).

10.2 Descriptive epidemiology, i.e. case finding: the prevalence and distribution of identified cases

10.2.1 Monitoring

The initial objective of the NHA-2 was to repeat the NHA-1 of 1983 (Hodiamont et al. 1987) in order to measure the psychiatric prevalence rate and its distribution in the general population in 1997 (descriptive epidemiology) and to compare these findings with those of the NHA-1. In the NHA-1 it was decided to assess psychiatric morbidity in the clinical tradition, i.e. at that time with the PSE-9 (Hodiamont 1986). In the NHA-2, we used the PSE-10, the principal part of the SCAN-2.1, which is the successor to the PSE-9, as the diagnostic instrument of choice in order to compare the results of the two surveys.

Both clinical, semi-structured interviews, i.e. the PSE-9 and the SCAN-2.1, need to be administered by clinically experienced interviewers. Notwithstanding their common roots, there are noteworthy differences between the PSE-9 and the SCAN-2.1 (see Table 10.2.1).

Because of the 23-year time gap between the introductions of the respective interviews, the then current classification system was used for each interview, with the possibility of an extended diagnostic range for the SCAN-2.1.

The number of items increased from 140 in the PSE-9 to 1899 in the June 1997 version of the SCAN-2.1. This substantial augmentation was a result of several developments. First, in order to fulfil the requirement of a separate examination of all possible signs and symptoms, many complex items of the PSE-9 had to be broken up. Second, two major classification systems are currently in use (DSM-IV and ICD-10). As there are some differences in the criteria used in the two systems, the number of items in the SCAN-2.1 was extended to cover all the data that have to be gathered so that these data can be classified in accordance with the two systems. Finally, as the number of diagnoses included in the classification systems has substantially increased, the number of items in the SCAN-2.1 had to be increased so that the diagnostic range would be covered.

The difference in rating scales between the PSE-9 and the SCAN-2.1 mainly concerned the additional sub-clinical rating possibility (“1”) in SCAN-2.1, which enabled a more refined registration of all signs and symptoms, especially in general population surveys (Giel & Nienhuis 1996). Finally, the algorithm of the PSE-9 was not directly accessible and could be used to calculate a diagnosis only when the severity of the symptoms reached a threshold. In contrast with its predecessor, the algorithm of the SCAN-2.1 is transparent and diagnoses are directly calculated on the basis of the fulfilment of DSM-IV criteria.

With respect to this monitoring study, the differences between the two interviews listed in Table 10.2.1 were resolved by converting PSE-10/SCAN scores into PSE-9 scores.

Table 10.2.1
Differences between the PSE-9 and the PSE-10 (SCAN-2.1) interview

	PSE-9	PSE-10/SCAN-2.1[xviii]
Year of introduction	1974	1997
Classification	ICD-9	DSM-IV/ICD-10
Diagnostic range	schizophrenic psychoses paranoid states manic and mixed affective depressive psychoses inhibited depressions neurotic depressions anxiety neuroses	psycho-organic disorders substance-related psychotic disorders mood disorders anxiety disorders eating disorders somatoform disorders dissociation disorders sleep disorders other psychiatric disorders
Number of items	140	1899
Possible scores	0 symptom absent 1 symptom moderate 2 symptom severe	0 symptom absent 1 symptom sub-clinical 2 symptom moderate 3 symptom severe
Algorithm	black box, indirect	open system, direct

10.2.1.1 Monitoring outcomes

10.2.1.1.1 Prevalence rates

The most conspicuous finding is that the prevalence rate for caseness increased by as much as 50% in the 14-year interval (1983-1997) in the Nijmegen Health Area (7.8% in 1983 and 11.9% in 1997).
The neurotic disorders (depression and anxiety disorders) contributed greatly to the prevalence rate for caseness (7.7% in 1983 and 9.7% in 1997). Although psychotic disorders accounted for a relatively small augmentation of the overall prevalence rate, they increased more than twofold over the 14-year period (0.8% in 1983 and 1.7% in 1997), which is substantially more than the almost one quarter increase for the neurotic disorders.

10.2.1.1.2 Distribution

We found no shift in the socio-demographic distribution of the disorders. The finding in the NHA-1 that females and males had equal prevalence rates was an exception to the usual pattern of distribution at that time, in particular a higher

prevalence rate for females than for males (Dohrenwend & Dohrenwend 1974; Surtees et al. 1983). This finding was not substantiated by the NHA-2 data, for, as expected, the prevalence rate was higher for females than for males. Consequently, the explanation that “role patterns in the Nijmegen Health Area may have changed in such a way that the inherent stress produced a similar amount of psychiatric disorder for the two sexes” (Hodiamont et al. 1987) can no longer be considered tenable.

10.2.1.2 Discussion of monitoring outcomes

When interpreting the monitoring outcomes in search of possible explanations of the shifts found, we have to take not only individual, but also societal and community factors into account, as the latter two determinants shape the experiences of persons living in these groupings. As for the societal and community levels, there were five notable observations.

First, consistent with the large increase in the prevalence rate in our monitoring study, in a large-scale national Dutch replication study in GP practices, Verhaak et al. (2005) reported a substantial increase in neurotic problems over a 14-year interval (1987 - 2001). This study revealed that the prevalence rate of the GHQ-12 scores ≥ 2 had risen from 16.8% in 1987 to 22.8% in 2001, which amounted to an increase of 36%.

Second, the prevalence rate of psychotic disorders had doubled over the 14-year interval. This increase can be explained by the planned de-institutionalisation in the Netherlands in the late 1980s and early 1990s. For psychotic patients, the reorientation of inpatient care from long stay to acute or short stay care resulted in a revolving-door phenomenon, in which a minority of the former psychotic long-stay inpatients accounted for a high percentage of short-stay re-admissions (Ooms et al. 2002). Consequently, a larger number of psychotic patients lived in the open population and were potential subjects for the survey sample of 1997.

A third circumstance was a complex of socio-demographic shifts: the increase in the mean highest (completed) level of education; the growing percentage of persons remunerated for their work; a decrease in the percentage of housewives; the substantial shift of persons moving to newly built areas in villages in the surrounding communities (SC). As an example of a societal factor, in this case the migrating city dwellers might have brought their neurotic burden with them, thereby contributing to a higher prevalence rate in the SC. In addition, after moving to the newly built areas, they might have lost their social support network and social cohesion, which had been beneficial to them before as a concomitant of city life. This unfortunate combination is in line with the thinking of Van Os (2000), who, in his article on the neighbourhood variation in the incidence of schizophrenia, concluded that the neighbourhood environment

modifies the individual risk for the disease and that premorbid vulnerability may be more likely to progress to overt disease in an environment with a higher perceived level of social isolation.
Fourth, there is a paradox between the growth in prosperity of the population as a whole and the significant increase in psychiatric prevalence rates. Hodiamont et al. (2005) reported an overall increase in GHQ scores over the 14-year interval and, in particular, a deterioration of the scores for the GHQ items relating to social interaction, while the scores for the GHQ items on socio-economic prospects were unchanged or even better. Verhaak et al. (2005) reported the same findings in their national Dutch replication study on GP practices (1987 - 2001). They found an increase in family and relational problems, while material problems were less often reported. As an illustration, in our study we found that in 1997 both students and housewives participated more in the workforce than they did in 1983, which brought about socio-economic prosperity, but also led to two or more (sometimes confusing or conflicting) simultaneous social roles and consequently to greater stress. Concomitant with this work role, the time for supportive informal social interaction diminished. Despite the socio-economic prosperity, this social isolation might be regarded as a major environmental consequence, associated with higher stress levels and a lesser feeling of well-being.
Finally, the fifth circumstance concerns the impending inundation of some of the villages of the SC. Part of the increase, especially in the neurotic (anxiety) states, can be explained by the fact that 1½ and 3½ years before the 1997 survey, the part of the NHA near the large rivers was evacuated because of the threat of a major inundation. As a possible result, the prevalence rate of neurotic disorders (mainly mood and anxiety disorders) in these threatened areas was about 25% higher than in the other villages of the SC.
This repeated cross-sectional survey provided us with a unique opportunity to explore time trends. We found neither age nor cohort effects. All five above-mentioned circumstances or factors reflected period effects, which together impacted heavily on the psychiatric prevalence rates.

10.2.2 Mapping

In the NHA-2 mapping study, the complete diagnostic range of the DSM-IV was used for the first time. This innovation made it possible to investigate the impact of the extension of the diagnostic range.
The following are major sources of variance that underlie the differences in prevalence rates found between studies – sources that are not based on real variations in the population: the sampling strategy, the age range, the classification systems, the diagnostic range, the time reference frame, the

interviewer's technical skills, the setting of the interview, the interview strategies and the use of technological aids (Fontenelle et al. 2006). In the comparison of the mapping results with those of other surveys, these sources of variance were taken into consideration.

10.2.2.1 Mapping outcomes

10.2.2.1.1 Prevalence

In the NHA-2, we found an overall prevalence rate of the integral range of psychiatric disorders in the general population of 17.6%. With respect to diagnostic categories, the highest prevalence rates were found for sleep, anxiety, substance use-related and mood disorders (4.8%, 4.6%, 4.0% and 3.8%). The three psychiatric diagnostic categories newly determined in the general population contributed substantially to the prevalence rate, which increased by almost fifty percent when they were included (11.9% - 17.6%). The cumulative prevalence rate of the three categories (9%) contributed (+5.7%) not only to the overall prevalence rate, but also, and importantly, to the comorbidity rate.

The comparison of the NHA-1 with the NHA-2 study also made explicit the influence of the diagnostic range. The prevalence rate when the PSE-9 design was used was 11.9% (monitoring [ICD-9], see chapter 5), while the rate was 17.6% (mapping [DSM-IV], see chapter 6) when the SCAN (PSE-10) design was used.

10.2.2.1.2 Distribution

In the NHA-2, the highest prevalence rates were found amongst females, city dwellers, the unmarried, and the underprivileged groups (the unemployed, chronically ill and the lowest educated).

With respect to the three newly determined categories, the prevalence rates for somatoform disorders and sleep disorders were higher in the SC than in Nijmegen, which contrasts with the distribution of higher prevalence rates that is usually found in the more urbanised regions. This deviation disappeared when the hierarchical rules were omitted (see Table 10.2.2) since these categories were then able to emerge.

Table 10.2.2
Distribution of somatoform disorders and sleep disorders across urbanisation, with and without hierarchy

	With hierarchy		Without hierarchy	
	Nijmegen	SC	Nijmegen	SC
Somatoform disorders	2.4	2.9	3.9	3.4
Sleep disorders	4.5	4.8	6.8	5.8

10.2.2.1.3 Comorbidity

Psychiatric comorbidity is a major health issue, given that it is associated with an extensive use of the health services and a substantial economic cost to society. Furthermore, most interventions (medication, psychotherapy, etc.) are not 'comorbidity proven', in that their effectiveness has been shown only for the treatment of well-defined monomorbid disorders.

In the NHA-2 study, the comorbidity rate was substantial, with 26% of the subjects identified as a case.

The additional three newly determined categories (dissociative disorders, somatoform disorders and sleep disorders) contributed to the comorbidity rate (36.7%) to a greater extent than the commonly determined part of the diagnostic range (22.3%). As these categories are partially hidden as a result of the DSM hierarchical rules, comorbidity would be much higher still if these rules were not taken into account.

10.2.2.2 Discussion of mapping outcomes

10.2.2.2.1 Prevalence rate

To our knowledge, the NHA-2 was the first psychiatric survey worldwide in which the prevalence rate for the full diagnostic range was estimated with a clinical, semi-structured, diagnostic instrument. Although the full range of diagnostic categories was used, the 1-month prevalence rate was moderate.

The number of studies that are in some respect methodologically comparable with the NHA-2 are limited. All comparisons fall short in one way or another with reference to either the use of a different diagnostic interview, a different (version of the) classification system, a different age range or a different range of the diagnoses that were assessed.

Nevertheless, some recent surveys are worthy of mention since they create a reference frame with regard to the NHA-2 mapping study (see Table 10.2.3).

Prevalence rate: SCAN studies

A comparison of studies conducted with the SCAN is hindered by the fact that the SCAN is rarely used in open population studies, and when it is, an earlier version of the SCAN is used, in which classification is done with the ICD-10. Furthermore, limitations exist with respect to the generalisability of the populations of these studies, the diagnostic range used and the variability of age ranges.

In the Formentera study, which was performed on a small Balearic Island on which educational standards were low (Roca et al. 1999), the SCAN-1 and the ICD-10 classification were used with a smaller diagnostic range than in the NHA-2. A 1-month prevalence rate of psychiatric disorders of 21.4% was found. In the Derry County study (McConnell et al. 2002) the SCAN-1 and the ICD-10 were used in a rural area of Northern Ireland with a high unemployment rate. The diagnostic range and the age range were limited, resulting in a 1-month prevalence rate of mood disorders, anxiety disorders and substance-related disorders of 7.5%. In the Camberwell Needs for Care Survey in an inner city area characterised by substantial social deprivation and a high proportion of ethnic minorities (Bebbington et al. 1997), the SCAN-1 and the ICD-10 classification were used with a smaller diagnostic range than in the NHA-2. The 1-month prevalence rate of psychiatric disorders was 9.8%.

When the diagnostic range of the NHA-2 was decreased to the size of the range used in the Formentera study (21.4%), the NHA-2 prevalence rate dropped from 17.6% to 11.7%. However, when, in addition, the generally low educational level of the subjects in the Formentera study was used as criterion for selecting a sub-sample of the NHA-2, the difference between the two studies was substantially smaller (NHA-2 16.8%). With reference to the Derry County study (7.5%), when the results of the NHA-2 mapping study were recalculated on the basis of the same limited diagnostic range for the rural area only, the prevalence rate was 10.9%. If the diagnostic range had been limited to the diagnostic categories adopted in the Camberwell study (9.8%) and the smaller age range and the inner city area were taken into account, the NHA-2 prevalence would have been 10.2%.

The difference in 1997 between the results of the monitoring (see chapter 5, PSE-9 prevalence rate 11.9%) and the mapping study (see chapter 6, SCAN/PSE-10 prevalence rate 17.6%) can be accounted for by the difference in age range and diagnostic range between the two studies. When we corrected the SCAN data of 1997 for the smaller age-range[xix] and limited the diagnostic range to the major diagnostic categories[xx], the prevalence rate for SCAN/PSE-10 was 11.9%, which was the same as the PSE-9 prevalence rate.

With respect to psychotic disorders, the ICD-9 classification system used in 1983, and consequently in the monitoring study, resulted in a substantial difference with reference to the DSM-IV system that was used for mapping in 1997. When the 1997 data were classified with the ICD-9 (as was done in the monitoring study), the prevalence rate of psychotic disorders (including affective psychoses) was 1.7%. When the 1997 data were classified with the DSM-IV, the prevalence rate of psychotic disorders was 0.5%, mainly because the affective psychoses were classified as mood disorders. This illustrates the importance of the use of a single well-accepted classification system for the interpretation of epidemiological data.
It can be concluded that under similar conditions, the prevalence rates of NHA-2 are quite similar to those of other SCAN-based studies, even though at first glance the results show major dissimilarities.

Prevalence: CIDI studies
In general, fully structured, non-clinical interviews yield higher diagnostic prevalence rates than clinical, semi-structured interviews (Alonso & Lepine 2007; Brugha et al. 1999a; Brugha et al. 2001; Cooper & Singh 2000; Goldberg & Huxley 1992; Kessler et al. 2005b). This disparity is possibly due to the measurement of the morbidity perceived by the subject, instead of the signs and symptoms validated by the clinician. The prevalence figures from Dutch CIDI studies range from 16.5% in the NEMESIS study (first wave) (Bijl et al. 1998) to 23.2% as determined with the CIDI in the Dutch capital Amsterdam shortly after the crash of an El-Al Boeing plane in the south-east part of the city (van Limbeek et al. 1994). The comparison of the NHA-2 with the NEMESIS first wave, described in chapter 6, was hampered by the use of the CIDI-2.1 and the DSM-III-R, the limited age range and the limited diagnostic range (Bijl et al. 1997b; Bijl et al. 1997a). When the diagnostic and the age range of the NHA-2 were reduced to the size of the ranges in the NEMESIS, the NHA-2 prevalence rate was 11.7%.
Narrow et al. (2002) introduced the clinical significance criterion[xxi] into the interpretation of the results of the NCS study (originally the CIDI-1.1 and the DSM-III-R were used with a limited diagnostic and age range (Kessler et al. 1994)) and as a result, the original 1-year prevalence rate decreased from 30.2% to 20.6%. In the replication study of the NCS (NCS-R), the CIDI-3.0 was used to assess 1-year prevalence rates (26.2%, see Table 10.2.3) on the basis of a limited diagnostic range (Kessler et al. 2005b).
In the European Study of the Epidemiology of Mental disorders (ESEMeD), the CIDI-3.0 and the DSM-IV were used in a large-scale general population survey in six European countries. For a limited diagnostic range in subjects aged 18

years or older, the researchers established the 1-year prevalence rate at 11.5% (Alonso et al. 2004; Alonso & Lepine 2007).

The general conclusion is that the only comparison under similar conditions that can be made between the NHA-2 and the CIDI studies is with the NEMESIS since a 1-month reference period was used for it, as it was for the NHA-2. Nevertheless, the prevalence rates reported from that study, in which a fully structured, non-clinical interview was used, were substantially higher than those of the NHA-2 with its semi-structured, clinically oriented interview. The comparison of the NHA-2 results with those of all other CIDI studies mostly shows even larger differences, probably due to a combination of the use of a fully structured, non-clinical interview with the 1-year assessment instead of the present state, i.e. 1-month prevalence rates.

Table 10.2.3
Prevalence rates of the NHA-2, Formentera , Derry County Study, the NEMESIS, the NCS (original, revised and replication) and the ESEMed

	NHA -2	Camberwell [a]	Formentera [b]	Derry [c]	NEMESIS [d]	NCS [e]		NCS-R [f]	ESEMeD [g]
	N = 1,799	N = 760	N = 697	N = 923	N = 7.076	N = 8,098		N =	N = 21,425
Interview instrument	SCAN-	SCAN-1	SCAN-1	SCAN-1	CIDI-2.1	CIDI-1.1	revised	CIDI-3.0	CIDI-3.0
Classification system	DSM-IV	ICD-10	ICD-10	ICD-10	DSM-III-R	DSM-III-R	DSM-III-R	DSM-IV	DSM-IV
Reference period	1-month	1-month	1-month	1-month	1-month	1-year	1-year	1-year	1-year
Age range (yrs)	18 – 75	18 – 65	15 – 93	18 – 65	18 – 65	18 – 54	18 – 54	≥ 18	≥ 18
Caseness	17.6	9.8	21.4	7.5	16.5	30.2	20.6	26.2	11.5
Psycho-organic disorder	0.3	---	---	---	---	---	---	---	---
Dementia	0.3	---	---	---	---	---	---	---	---
Other organic brain disorder	0.0	---	---	---	---	---	---	---	---
Substance-related disorder	4.0	---	6.0	1.6	5.8	11.5	7.6	3.8	1.0
Dependence	0.9	---	3.7	1.6	3.3	---	---	---	---
Abuse	3.2	---	---	---	2.6	---	---	---	---
Psychotic disorder	0.5	---	0.5	---	0.2	0.2	0.2	---	---
Schizophrenia	0.4	---	---	---	---	---	---	---	---
Delusionaldisorder,non-aff. psychosis	0.1	---	---	---	---	---	---	---	---
Mood disorder	3.8	3.3	3.4	2.4	3.9	11.1	7.5	9.5	4.5
Depression	2.7	3.1	1.6	2.4	2.7	10.1	6.4	6.7	4.1
Mania, hypomania, mixed disorder	0.3	0.1	0.6	---	0.6	1.5	1.5	2.6	---
Dysthymic disorder	1.9	0.4	3.1	---	1.6	2.5	1.8	1.5	1.1
Anxiety disorder	4.6	4.2	---	3.5	9.7	18.7	12.1	18.1	8.4
Generalised anxiety disorder	0.3	0.7	0.8	0.2	0.8	3.4	2.8	3.1	0.9
Panic disorder	0.3	0.5	0.3	2.4	1.5	2.2	1.7	2.7	0.7
Agoraphobia (without panic disorder)	0.2	0.7	---	0.6	1.0	3.7	2.2	0.8	0.3
Specific phobia	0.8	1.6	3.4	0.2	5.5	8.6	4.4	8.7	5.4
Social phobia	1.3	0.4	0.9	---	3.7	7.4	3.7	6.8	1.6
Obsessive-compulsive disorder	1.3	---	0.9	---	0.3	---	---	1.0	---
Eating disorder	0.4	---	1.9	---	0.3	---	---	---	---
Anorexia Nervosa	0.1	---	0.1	---	0.0	---	---	---	---
Other eating disorder	0.3	---	1.9	---	0.3	---	---	---	---
Somatoform disorder	2.8	---	---	---	---	---	---	---	---
Somatisation disorder	0.1	---	---	---	---	---	---	---	---
Pain disorder	1.2	---	---	---	---	---	---	---	---
Other somatoform disorder	1.5	---	---	---	---	---	---	---	---
Dissociation disorder	1.4	3.0	---	---	---	---	---	---	---
Sleep disorder	4.8	---	---	---	---	---	---	---	---
Dyssomnia	2.2	---	---	---	---	---	---	---	---
Parasomnia	2.1	---	---	---	---	---	---	---	---
Other sleep disorder	0.5	---	---	---	---	---	---	---	---

Legend to table 10.2.3
[a] (Bebbington et al. 1997); [b] (Roca et al. 1999); [c] (McConnell et al. 2002); [d] (Bijl et al. 1997a); [e] (Kessler et al. 1994) and (Narrow et al. 2002); [f](Kessler et al. 2005b); [g] (Alonso et al. 2004; Alonso & Lepine 2007)
* only alcohol-related disorders

10.2.2.2.2 Distribution

The socio-demographic distribution of the prevalence rates found in the NHA-2 agrees with those found in most other large-scale surveys (Alonso & Lepine 2007; Bijl & Ravelli 2000; Brugha et al. 2004; Dohrenwend & Dohrenwend 1974; Kessler et al. 2005b). Fryers et al. (2005) concluded that "people of lower socio-economic status, however measured, are disadvantaged, and this includes higher frequencies of the conditions now called the 'common mental disorders'[xxii]." All in all, the interview tradition does not affect the distribution of psychiatric disorders in the general population, which is consistent with the findings of Surtees et al. (1983).

10.2.2.2.3 Comorbidity

In the NHA-2 study, the comorbidity rate was 26% of all subjects identified as a case. A comparison of the comorbidity rates found in the NHA with those found in other studies comes up against the same difficulties as those mentioned above with respect to the prevalence rates. The only comparable general population study in which the SCAN was used as the diagnostic interview was the Formentera study by Roca et al. (1999), who reported a comorbidity rate of less than 27% for those subjects identified as cases. This rate is almost the same as that found in the NHA-2. As far as we know, no comorbidity rates were reported in the Derry County study.
The majority of the general population studies in which comorbidity rates were reported used a fully structured, non-clinical diagnostic interview and adopted different reference periods from those used in the NHA-2, in addition to differences in the diagnostic and the age range.
Only the NEMESIS study[xxiii] reported an overall 1-month comorbidity rate that was slightly higher (30%) than that found in the NHA-2 (Bijl et al. 1998; Ravelli et al. 1998). Other studies reported only 1-year and lifetime measures, with higher comorbidity rates. There are several possible reasons for these artificially higher comorbidity rates: the use of various definitions of comorbidity[xxiv], the variability of the classification systems[xxv] and the consequences of fully versus semi-structured interviews[xxvi].

Although we found a comorbidity rate that was more moderate than the rates determined in other studies, comorbidity is widespread in the general population. Comorbidity has serious consequences for the individual and is responsible for a heavy burden on society (Bijl et al. 1998; Kessler et al. 1997). Comorbidity also poses problems for therapists as most therapeutic interventions are based on so-called monomorbid evidence.

10.2.2.2.4 Three newly determined diagnostic categories

The three psychiatric diagnostic categories (dissociative disorders, somatoform disorders and sleep disorders) newly determined in the general population contributed substantially to the psychiatric prevalence rate, which increased by almost fifty percent (from 11.9% to 17.6%) when they were included. Bebbington et al. (1997) reported that the inclusion of sleep disorders resulted in a prevalence-rate increase of almost one third. The importance of these categories is illustrated by Janca et al., who stated that much research has recently been conducted on somatoform disorders and that these investigations have established that they are clinically important, being associated with a serious burden on health services and the economy (Janca et al. 2006).

Relative to the diagnostic categories on which measurements are usually carried out, these three additional categories turned out to occupy uncommon positions in the diagnostic system.

First, diagnoses of at least two categories (sleep and somatoform disorders) share several symptoms with other (major) diagnoses in the DSM-IV. Consequently, they are especially sensitive to the hierarchy rules of the DSM system.

Second, the application of the hierarchy rules might also explain the spurious lower prevalence rates for somatoform disorders and sleep disorders in Nijmegen with respect to the SC. The signs and symptoms of the two disorders are integrated into disorders like depression, which are higher in hierarchy and have a significantly higher prevalence rate in the city than in the SC.

Third, the prevalence rates of somatoform and sleep disorders were significantly affected by the way in which they were assessed; in particular, when clinical judgement was omitted, these prevalence rates increased substantially (see chapter 8). This finding is supported by those of other studies (Jacobi et al. 2004; Quintana et al. 2007).

Fourth, the three categories contributed to comorbidity to an even greater extent than the normally assessed diagnostic categories. Given their high prevalence rate, among other factors, they feature prominently in the overall mental health status.

Currently, in the light of the development of the DSM-V, the position of somatoform disorders is being called into question (de Waal et al. 2006; Henningsen & Lowe 2006; Mayou et al. 2005; Rief et al. 2006; Starcevic 2006; Stein 2006; Waller & Scheidt 2006). These authors differ in the solutions they propose for the problem of an overly inclusive conception of the category of somatisation disorders. Stein suggests limiting this category to only the few severely dysfunctional patients for whom it is so painful to acknowledge feelings that affects are expressed through physical symptoms. Mayou et al. take an extreme standpoint and transfer most of the somatoform disorders to the (physical) axis III of the classification system, as a useful step toward the elimination of unhelpful dualist thinking. On the other hand, Rief and Starcevic argue that since somatisation is clinically important and associated with a burden on the health services and the economy, the political pressure to change the category should be resisted, and it should be better defined rather than abolished.

10.2.3 Conclusions regarding descriptive epidemiology

Many of the sources of variance mentioned by Fontenelle et al. (2006) still play an important role in descriptive epidemiology. All these sources complicate simple comparisons of gross prevalence rates. In particular, classification systems and interview strategies are important as they tend to disappear into the background and become implicit as soon as prevalence rates are calculated and reported.

With respect to the first objective of this thesis (comparison of the NHA-1 with the NHA-2 data), the monitoring of case finding revealed important shifts in the overall prevalence rates over the 14-year interval from 1983 to 1997 and a number of explanations have been advanced for these shifts. The semi-structured, clinically oriented interview SCAN-2.1 (DSM-IV) has shown consistency with the prevalence rates as calculated with the PSE-9 (ICD-9). Furthermore, it has been established that the SCAN-2.1 is suitable for general population surveys.

With respect to the second objective (to determine the consequences of the relative extension of the diagnostic range), the mapping, or case finding, revealed that the addition of the three newly determined categories resulted in a substantial contribution to the overall prevalence rate.

The various prevalence rates obtained in surveys in which the SCAN interview was conducted were fairly similar when the study conditions were made comparable. Prevalence rates obtained in surveys in which the CIDI was used were all higher than the aforementioned SCAN prevalence rates.

When the prevalence rates of diagnoses of a lower hierarchical ranking are discussed, they should be presented with and without the application of hierarchy rules so that the impact of these diagnoses on the total psychiatric burden can be understood.

Concerning the disorders of the three newly determined categories, a question that needs to be answered is whether they should be regarded as separate entities in the classification system. When a disease entity is not sequentially related to a disorder of a higher hierarchical ranking and when it is accompanied by a sufficient burden related to the overall health status, in our opinion, it deserves a separate place in the diagnostic range. When a disease entity is clearly sequentially related to a disorder of a higher hierarchical ranking, it should be included in the classification spectrum of that disorder, but, when a disease entity is not sequentially related to a disorder of a higher hierarchical ranking and is not accompanied by a sufficient burden to the overall health status, it deserves no place in the classification system at all.

10.3 Diagnostics: case identification

10.3.1 Psychometric properties

10.3.1.1 Psychometric outcomes

The psychometric properties of the SCAN-2.1 were tested in a stratified sample of the general population[xxvii] with two designs: a test-retest procedure and a standardised procedure. The test-retest procedure (two SCAN interviews administered to one subject by different interviewers within 1 week of each other) showed fair to moderate reliability. The standardised procedure (rescoring videotaped SCAN interviews, administered and scored for reference diagnoses by highly experienced clinicians) showed substantial to almost perfect agreement.

10.3.1.2 Discussion of psychometric outcomes

Our study differs from other psychometric studies in two ways: to begin with, it is the first study to explore the reliability and validity of the SCAN-2.1 with respect to the integral diagnostic range; second, for the purposes of this exploration, a stratified sample was used that reflected the general population. All former investigations into the psychometric properties of previous versions of the SCAN concerned specific sections tested in more or less clinically identified samples (Andrews et al. 1995; Easton et al. 1997; Tomov & Nikolov 1990; Wing et al. 1990).

A number of studies produced evidence of a lower reliability for the clinically oriented interviews than for the fully structured, non-clinical interviews (Andrews et al. 1995; Compton et al. 1996; Regier 2000; Wittchen 1994; Wittchen et al. 1999). On the other hand, there is also evidence that the SCAN-1, at least for certain diagnostic categories, is as reliable as the CIDI (Üstün et al. 1997) and that the CIDI-auto has poor validity with respect to clinical diagnoses (Rosenman et al. 1997).

There is abundant evidence in the literature that lay interviewers perform differently from clinicians when administering a diagnostic interview (Andrews et al. 1995; Anthony et al. 1985; Haro et al. 2006; Helzer et al. 1985; McLeod et al. 1990; Slade & Andrews 2002)[xxviii]. The most important difference seems to arise from the use of clinical judgement (see sections 10.3.2 and 10.3.3). Clinical judgement as used in the SCAN-2.1 results in an assessment of disease rates, whereas lay interviews generate rates on the basis of perceived illness.

In view of the preceding results and reflections, it can be concluded that the reliability and validity of the SCAN-2.1 is satisfactory for the assessment of psychiatric disorders in the general population.

Notwithstanding this general conclusion, some recommendations can be made on improving the use of the SCAN-2.1. First, some diagnoses depend completely on items for which there are explicit questions at the beginning of the section concerned. The answers corresponding to these items have to be re-evaluated later on in the interview without the aid of explicit questions. This routing of these questions is confusing and can lead to a loss of information. Consequently, during the 1-week training course, special attention should be given to intensively teaching the structure of the SCAN and its underlying sections. It might be beneficial if a checklist of items to be reviewed for completeness at the end of the interview were used or if the information were reappraised at the time at which it was obtained.
Second, from the standpoint of cost-effectiveness in general population studies, it might be more efficient to train all interviewers intensively on part one of the SCAN, and to select a few of them for training on the entire SCAN interview. If the screening section of psychotic and cognitive pathology (section 14) turns out to be positive, the clinically more experienced interviewer can conduct the second part of the SCAN interview.
Finally, thanks to the comprehensive SCAN training undergone by the interviewers, the instrument has proved its reliability and validity for use in research. Ahead of the next topics in sections 10.3.2 and 10.3.3, we are convinced that this conclusion applies to clinical practice too. In future research, exploration of the effects of clinical judgement with the full version of the SCAN-2.1 should be one of the principle objectives[xxix].

10.3.2 Clinical judgement

10.3.2.1 Clinical judgement outcomes

The results presented in chapter 8 show that clinical judgement plays a crucial role in the assessment of multiple diagnoses and consequently, in general, has an important influence on prevalence rates. The following results are noteworthy since they reflected major shifts with regard to clinical judgement.
First, not all diagnoses were equally affected by clinical judgement as operationalised in the SCAN. Diagnoses based on subjective experiences were more impacted than those based on behaviour. Second, some diagnoses (like somatoform disorder NOS) are extremely sensitive to clinical judgement since the presence or absence of a single sign or symptom (related to daily life problems) is crucial. Third, when clinical judgement was omitted, the diagnoses

of some of the subjects already identified as a case tended to shift towards the major diagnostic categories. Fourth, the prevalence rate of caseness increased by 25% when clinical judgement in the SCAN was omitted.

10.3.2.2 Discussion of clinical judgement outcomes

For the first time, thanks to its transparency we were able to explore the algorithm black box of a standardised interview (Kendell 1993). Clinical interviews, like the SCAN-2.1, combine the structure of the interview with the clinical judgement of the interviewer for the assessment of the clinical significance of the perceived signs and symptoms.

Our results show that clinical judgement gives rise to outcomes that are crucially different from the outcomes of interviews in which this method of decision-making is not used.

Why is clinical judgement important? Because clinicians start from the assumption that psychiatric morbidity more or less affects the patient's evaluation of his or her underlying signs and symptoms. Clinical judgement expresses a considered clinical opinion about which signs and symptoms, and thoughts and experiences can be regarded as part of normal life (sometimes in difficult circumstances) and which should be declared pathological. With respect to epidemiological studies, this statement is summarised by Cooper (1979), who argued that researchers should choose their (diagnostic) variables on the basis not only of scientific, but also of clinical relevance.

Clinical judgement is not merely a matter of adjusting the algorithm of fully structured, non-clinical interviews, but, as with the SCAN-2.1, it implies an explicit consideration preceding the acknowledgement of each sign and symptom. However, even when a clinical relevance criterion was introduced afterwards, as was done with the NCS revision, the prevalence rates were decreased by almost one third (Narrow et al. 2002). This finding stresses the importance of a differentiation between disease and illness measures and of the clarification of an adequate clinical judgement in diagnostic decision-making. The SCAN-2.1 has the properties needed to explore both these qualities. Even when no illness is manifest, a positive score for a sign or symptom can be obtained by the interviewer thanks to the opportunity that clinical judgement provides.

Consequently, it is imperative that more specific research work should be done on the process, the meaning and the impact of clinical judgement in the diagnostic process in psychiatry. Furthermore, we plead for the abolition of the formal routine of omitting the second part of sections in which core questions do not yield positive scores. Such action would contribute to a closer-to-complete picture of all signs and symptoms present in the general population and hence to a better understanding of the process of clinical judgement.

10.3.3 Comparison of the CIDI and the SCAN

10.3.3.1 CIDI-SCAN comparison outcomes

In the comparison of the CIDI with the SCAN, both interviews were administered to the same subjects of a stratified sample of the open population. Consequently, the focus was not on the actual prevalence rates but on the comparison of the results generated by the two instruments. And striking differences between the two sets of results were indeed found. First, there were seven strong canonical correlations (0.85 - 0.62) between the set of depression-related CIDI items and the set of depression-related SCAN items. The explained percentual variance of four of the CIDI correlates was substantial (0.11 - 0.13) and this was also the case for the SCAN (0.07- 0.09). Second, and paradoxically, the kappa (κ) values with respect to the agreement between the individual criteria of the depressive episode yielded by the CIDI and the SCAN were modest (0.03 - 0.53). Third, the overall κ values with respect to the mutual recognition of caseness (0.46) and diagnostic categories (0.09 - 0.51) were modest as well.

10.3.3.2 Discussion of CIDI-SCAN comparison outcomes

There are a limited number of direct comparative studies on the assessment of the CIDI and the SCAN. Most of these concentrate on a single diagnostic DSM-IV category and all report disappointing agreement values, with the odd encouraging exception (Andrews et al. 1995; Andrews & Peters 1998; Compton et al. 1996; Pull et al. 1997; Shibre et al. 2002).

At first glance, the CIDI and the SCAN results of the NHA-2 disagree substantially. But an in-depth analysis (canonical correlation) of the latent structure of the signs and symptoms of the depressive episode (DSM-IV) discloses an appreciable concordance between the two instruments. The canonical correlations are high and the underlying symptomatology (with reference to the SCAN and the CIDI items) shows high loadings on the contributing latent variables, although the latter are composed of a mix of single items. The explained variance of the first seven correlates with respect to both the CIDI and the SCAN is considerable (62% and 46%). Both interviews appear to measure the same burden, but each in its own way, with divergent operationalisations, which are reflected by appreciable differences in the prevalence rates found in studies in which one or other of these interviews is used.

Most prevalence studies yield a lower rate for clinical, semi-structured as opposed to fully structured, non-clinical interviews (Alonso & Lepine 2007; Brugha et al. 1999a; Brugha et al. 2001; Cooper & Singh 2000; Goldberg &

Huxley 1992; Kessler et al. 2005b). The CIDI tends to generate more diagnoses than clinicians claim are clinically relevant (Andrews & Peters 1998; Rodgers & Mann 1986).

Andrews et al. (1995) emphasised that “the less than perfect concordance between the CIDI and the SCAN is unlikely to be due to differing definitions of diagnoses”. Although we found high correlations between clusters of depression items on the SCAN and the CIDI, agreement between the individual depression items was poor. This was due to the fact that it proved difficult to match the corresponding items of the SCAN and the CIDI that represent the operationalisation of the DSM-IV diagnostic criteria for mood disorders. This divergence with reference to item operationalisation causes a fundamental mismatch between the two instruments, a mismatch that is amplified by the differences in the procedures for obtaining diagnoses. Consequently, we can not agree with the standpoint of Andrews et al. that “the similarity in the weights within the canonical variates suggests that the instruments produce similar diagnostic discriminations”.

10.3.4 Conclusions regarding diagnostics

The semi-structured SCAN-2.1 interview, with its clinical judgement, is suitable for case identification in the general population. The use of clinical judgement has important effects on case identification. These effects manifest themselves in a lower overall prevalence rate and a category-dependent lower prevalence rate for specific diagnoses. The CIDI and the SCAN (the major representatives of the two psychiatric diagnostic interview traditions) concord strongly with respect to a general level of depressive burden, but the agreement is poor to moderate for individual depression items, diagnostic criteria and psychiatric categories.

Psychiatric epidemiological population studies should be based on clinically relevant variables, so that clinical judgement is an indispensable instrument for harvesting valid information on diseases. The availability of semi-structured, clinically oriented instruments like the SCAN-2.1 ensures that such data can be captured. The SCAN-2.1 yields disease rates as opposed to the illness rates found with fully structured, non-clinical instruments like the CIDI. Most evidence indicates that the CIDI is somewhat more reliable than the SCAN, but in our view, a maximal reliability rate should not be achieved at the expense of clinical validity. We are therefore convinced that clinically oriented instruments should be used, even though to some extent, this practice may incur a slight loss of reliability.

Many of the sources of variance mentioned by Fontenelle et al. (2006) still play a role in the diagnostic process, but it is difficult to quantify their impact on case identification.

Notwithstanding, classification systems and interview strategies are important, as they tend to recede to the background and become implicit as soon as prevalence rates are calculated and reported.

10.4 General conclusions and recommendations

In retrospect, several topics stand out in this thesis that we will proceed to review briefly under the headings general conclusions, study limitations, recommendations and final remark.

10.4.1 General conclusions

As mentioned in the research questions (see chapter 4.4), this thesis uses descriptive epidemiological data generated by the SCAN-2.1 with the objective of focusing on the relevance of clinical judgement in psychiatric diagnostics.

With respect to the descriptive epidemiological data, we found that the prevalence rate of psychiatric morbidity in the open population increased by almost fifty percent over the 14-year interval from 1983 to 1997. In the NHA-2, the distribution of psychiatric morbidity reflected a well-known pattern and the expansion of the diagnostic range brought about a significant increase in the overall prevalence rate and the comorbidity rate.

The SCAN-2.1 proved to be a reliable and valid instrument and one that is sensitive to changes over time.

We used descriptive epidemiological data to explore the sub-clinical rating “1” and found that the operationalisation of clinical judgement in the SCAN-2.1 differentiates many of the presented signs and symptoms as sub-clinical, thereby decreasing the potential prevalence rates. To compare the SCAN-2.1 with the CIDI we used a stratified sample of the general population and found that although a similar burden was assessed with the two instruments, they diverged substantially in their operationalisations, these divergences having manifested themselves in different prevalence rates.

10.4.2 Study limitations

The first point to note is that the cross-sectional design of the NHA studies contained potential limitations with regard to the comparability of the study populations and the compatibility of the measuring instruments and the diagnostic systems used. We controlled for these disadvantages and feel confident that we minimised the potential effects of this limitation. On the other hand, a cross-sectional design has the advantage that it makes possible an examination of secular population changes.

Second, the non-response in 1997 appears to call into question the representativeness of the survey sample. Potential subjects seemed to be deterred by the exhaustive information campaign deemed necessary for informed consent by the medical ethics committee. The sampling design, in

which GP practices were used, however, enabled us to check for selectivity by means of the prescription data on psychotropic medication. There was no evidence of a selection bias in the sense of patients with (medicated) psychiatric morbidity being overrepresented in the sample. Furthermore, we corrected statistically for over- and under-response with respect to socio-demographic variables.
Third, the choice of an expensive and demanding clinical interview imposed limitations on the number of interviewees and thus on the power of the study. Nevertheless, our findings show significance and clinical relevancy.
Finally, the data gathered for this study were restricted to those necessary to attain the primary objective of the NHA-2 project, which was to carry out an epidemiological survey. Consequently, we were not able to operationalise all forms of clinical judgement. Consequently, the differences between prevalence rates with and without clinical judgement probably are somewhat underestimated.

10.4.3 Recommendations

On the basis of the reflections and conclusions set out in the previous paragraphs, we shall now formulate some recommendations, which we will group under five themes:

Clinical judgement

In this study, in which it was possible to register the sub-clinical status of assessed signs and symptoms (illness, but not disease), we established that clinical judgement is of major importance and has vital consequences for the process of case identification.

a. Replication studies with respect to the preceding finding are advisable. For these studies, it is recommended to omit the cut-off points in the routing of the SCAN-2.1 interview in order to assess all signs and symptoms, including those enquired about in the second part of sections in which core questions do not yield positive scores.
b. In addition, the situation in which the subject does not present illness but the interviewer decides to classify a sign or symptom as present (disease or sub-clinical presence) needs to be investigated more explicitly. Such studies must provide for the registration of the subject's verbatim answers.
c. Similarly, it should be determined which signs and symptoms can be impacted by one or more types of clinical judgement and which can not be. On the one hand, this information can facilitate the search for less sensitive equivalents of those signs and symptoms that are (too) sensitive to clinical judgement, and on the other hand, the information can be used to identify sensitive signs and symptoms that should not be assessed with a non-

clinical interview. Such studies would facilitate the comprehension of the underlying mechanisms of clinical judgement, thereby helping us to specify, standardise, develop and simplify clinical judgement.

d. The reliability and validity of all modalities of clinical judgement should be established explicitly.

The three additional categories

The right to exist of the three newly determined additional categories (dissociative disorders, somatoform disorders and sleep disorders) is supported by their relatively high prevalence rate (9%) in the open population.

a. The burden of subjects with a monomorbid manifestation of a disorder that is included in one of the three additional categories should be established and compared in a cross-sectional design with the burden of those subjects diagnosed with a monomorbid disorder that falls within the major diagnostic categories. These results would constitute evidence as to whether the additional categories were sufficiently important to occupy a position in the classification system in their own right. As some of the disorders have been found to be very sensitive to clinical judgement, a semi-structured, clinical approach would be needed.
b. Comorbidity data should be gathered to clarify the hierarchical position of the three additional categories in the classification system.
c. Likewise, the relation of the disorders included in the three additional categories to other correlates (for example coping styles, life events and social support) could provide information on the causes and effects of these disorders and such information would benefit curative and preventive interventions.
d. Longitudinal studies are needed to clarify the relations of the disorders included in the three additional categories to the other diagnoses: These disorders may be precursors of other, major disorders, separate disease entities or initial manifestations of complex syndromes which could persist for the rest of the subject's life.

Comorbidity

Whatever diagnostic interview is used, comorbidity is always of major importance since curative and preventive interventions typically are based on monomorbid disease entities.

a. To avoid the confusion brought about by the ambiguous use of the term comorbidity, whereby it refers not only to concurrent but also to consecutive pathologies, this term should be reserved for the situation in which disease entities occur at the same time, i.e. in which they are concurrent. The term consecutive morbidity should be used when two or more disease entities exist in a subject, but with a time gap between them.

b. Longitudinal research will show a picture not only of comorbidity (concurrent morbidity) but also of separate, consecutive morbidity. Such a view could teach us to differentiate between real and artificial comorbidity, provide information on possible precursor syndromes and provide clues on the prevention of a deterioration into a major psychiatric disorder.
c. Further study of the occurrence and composition of comorbidity as assessed with the SCAN-2.1 is needed since the SCAN will generate a mix of comorbid disease entities different from the mix that other instruments would disclose, among other reasons because clinical judgement is incorporated into the SCAN. Such data would enable us to investigate true comorbidity, its prevalence rate and distribution, and its importance for daily psychiatric practice.

The CIDI-SCAN comparison

It is unacceptable that the two major standardised interview traditions in psychiatry create divergent bodies of evidence with respect to caseness and diagnoses.

a. Detailed studies on the differences between the CIDI and the SCAN should be carried out in order to obtain data on the information variance, the observation and interpretation variance, and the criterion variance. Such investigations could be carried out through a comparison of video-registered double interviews with respect to the answers to the questions and the additional information obtained, and also of the ways in which the items registered are processed by the algorithms.
b. Such a study could contribute to the identification of those items for which clinical judgement is imperative, those for which it is advisable and those on which it would have no impact.

The SCAN

In the course of the NHA-2 project, a total of 983 subjects from the general population were interviewed with the SCAN-2.1, a semi-structured, clinically oriented instrument. To our knowledge, this was the first time that the SCAN-2.1 had been used for the assessment of psychiatric morbidity in the general population. We consequently consider it appropriate to make some observations on the use of the instrument itself.

a. The rating scales of the SCAN do not enable the registration of all forms of clinical judgement. Consequently, we propose that research should be done in which, in addition to the information gleaned with the formal SCAN, the verbatim answers of the subjects are registered.
b. For a real bottom-up procedure, the cut-off points after the core questions[xxx] in each section should be omitted so that a complete dataset can be obtained on subjects in the open population for all available items. Future

classification systems could be tested against these datasets for their innovative value.

c. Items without questions that are reviewed later on in the interview so that information for additional diagnoses[xxxi] can be registered should either be prompted for and accompanied by the information that was procured earlier in the interview, or else these items should be scored at the time the information is gathered.

10.4.4 Final remark

Despite the fact that the development of classification systems and the standardisation of diagnostic procedures have improved understanding in the psychiatric community, the results of epidemiologic research are still characterised by an abundance of data that admit of several interpretations. Although these developments are necessary, we conclude that they are not sufficient for a satisfactory contribution to epistemological unity. With reference to the interpretation and valuation of the captured data, currently the classification system and the interview strategy are kept hidden in the background. We strongly recommend that both the classification system and the interview strategy should be used as an *explicit* frame of reference when it comes to the valuation of the research results and that the conclusions from these results should be put into clinical perspective.

References

Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, de Girolamo G, Graaf R, Demyttenaere K, Gasquet I, Haro JM, Katz SJ, Kessler RC, Kovess V, Lepine JP, Ormel J, Polidori G, Russo LJ, Vilagut G, Almansa J, Arbabzadeh-Bouchez S, Autonell J, Bernal M, Buist-Bouwman MA, Codony M, Domingo-Salvany A, Ferrer M, Joo SS, Martinez-Alonso M, Matschinger H, Mazzi F, Morgan Z, Morosini P, Palacin C, Romera B, Taub N and Vollebergh WA. (2004) Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica Supplementum* 21-27.

Alonso J and Lepine JP. (2007) Overview of key data from the European Study of the Epidemiology of Mental Disorders (ESEMeD). *Journal of Clinical Psychiatry,* 68 Suppl 2, 3-9.

Andrews G and Peters L. (1998) The psychometric properties of the Composite International Diagnostic Interview. *Social Psychiatry and Psychiatric Epidemiology,* 33, 80-88.

Andrews G, Peters L, Guzman AM and Bird K. (1995) A comparison of two structured diagnostic interviews: CIDI and SCAN. *Australian and New Zealand Journal of Psychiatry,* 29, 124-132.

Anthony JC, Folstein M, Romanoski AJ, Von Korff MR, Nestadt GR, Chahal R, Merchant A, Brown CH, Shapiro S, Kramer M and . (1985) Comparison of the lay Diagnostic Interview Schedule and a standardized psychiatric diagnosis. Experience in eastern Baltimore. *Archives of General Psychiatry,* 42, 667-675.

Bebbington PE, Marsden L and Brewin CR. (1997) The need for psychiatric treatment in the general population: the Camberwell Needs for Care Survey. *Psychological Medicine,* 27, 821-834.

Bijl RV and Ravelli A. (2000) Psychiatric morbidity, service use, and need for care in the general population: results of The Netherlands Mental Health Survey and Incidence Study. *American Journal of Public Health,* 90, 602-607.

Bijl RV, Ravelli A and van Zessen G. (1998) Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Social Psychiatry and Psychiatric Epidemiology,* 33, 587-595.

Bijl RV, van Zessen G and Ravelli A. (1997a) Psychiatric morbidity among adults in The Netherlands: the NEMESIS- Study. II. Prevalence of psychiatric disorders. Netherlands Mental Health Survey and Incidence Study [in Dutch]. *Nederlands Tijdschrift voor Geneeskunde,* 141, 2453-2460.

Bijl RV, van Zessen G, Ravelli A, de Rijk C and Langendoen Y. (1997b) Psychiatric morbidity among adults in The Netherlands: the NEMESIS-Study. I. Objectives, design and methods. Netherlands Mental Health Survey and Incidence Study [in Dutch]. *Nederlands Tijdschrift voor Geneeskunde,* 141, 2248-2252.

Brugha T, Jenkins R, Bebbington P, Meltzer H, Lewis G and Farrell M. (2004) Risk factors and the prevalence of neurosis and psychosis in ethnic groups in Great Britain. *Social Psychiatry and Psychiatric Epidemiology,* 39, 939-946.

Brugha TS, Bebbington PE and Jenkins R. (1999a) A difference that matters: comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population [editorial]. *Psychological Medicine,* 29, 1013-1020.

Brugha TS, Jenkins R, Taub N, Meltzer H and Bebbington PE. (2001) A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychological Medicine,* 31, 1001-1013.

Brugha TS, Nienhuis F, Bagchi D, Smith J and Meltzer H. (1999b) The survey form of SCAN: the feasibility of using experienced lay survey interviewers to administer a semi-structured systematic clinical assessment of psychotic and non-psychotic disorders. *Psychological Medicine,* 29, 703-711.

Compton WM, Cottler LB, Dorsey KB, Spitznagel EL and Mager DE. (1996) Comparing assessments of DSM-IV substance dependence disorders using CIDI-SAM and SCAN. *Drug and Alcohol Dependence,* 41, 179-187.

Cooper B. (1979) *Demographic and epidemiological methods in psychiatric research*. Springer Verlag, Berlin.

Cooper B and Singh B. (2000) Population research and mental health policy. Bridging the gap [editorial]. *British Journal of Psychiatry,* 176, 407-411.

de Waal MW, Arnold IA, Eekhof JA and van Hemert AM. (2006) Somatoform disorders in general practice: prevalence, functional limitations and comorbidity with anxiety and depression [in Dutch]. *Nederlands Tijdschrift voor Geneeskunde,* 150, 671-676.

Dohrenwend BP and Dohrenwend BS. (1974) Social and cultural influences on psychopathology. *Annual Review of Psychology,* 25, 417-452.

Dohrenwend BP and Dohrenwend BS. (1982) Perspectives on the past and future of psychiatric epidemiology. *American Journal of Public Health,* 72, 1271-1279.

Easton C, Meza E, Mager D, Ulug B, Kilic C, Gogus A and Babor TF. (1997) Test-retest reliability of the alcohol and drug use disorder sections of the schedules for clinical assessment in neuropsychiatry (SCAN). *Drug and Alcohol Dependence,* 47, 187-194.

Eaton WW, Regier DA, Locke BZ and Taube CA. (1981) The Epidemiologic Catchment Area Program of the National Institute of Mental Health. *Public Health Reports,* 96, 319-325.

Feinstein AR. (1970) The pre-therapeutic classification of comorbidity in chronic disease. *Journal of Chronic Diseases,* 23, 455-468.

Fleming JA and Hsieh CC. (2002) Introduction to epidemiologic research methods. In MT Tsuang en M Tohen (Eds.), *Textbook in psychiatric epidemiology, second edition*. pp. 3-33. Wiley-Liss, New York.

Fontenelle LF, Mendlowicz MV and Versiani M. (2006) The descriptive epidemiology of obsessive-compulsive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry,* 30, 327-337.

Fryers T, Melzer D, Jenkins R and Brugha T. (2005) The distribution of the common mental disorders: social inequalities in Europe. *Clinical practice and epidemiology in mental health,* 1, 1-12.

Giel R and Nienhuis FJ. (1996) *SCAN-2.1: Schedules for clinical assessment in neuropsychiatry [in Dutch]*. WHO, Geneva / Amsterdam.

Goldberg DP and Huxley P. (1992) *Common mental disorders: a biosocial model*. Tavistock/Routledge, London.

Haro JM, Arbabzadeh-Bouchez S, Brugha TS, de Girolamo G, Guyer ME, Jin R, Lepine JP, Mazzi F, Reneses B, Vilagut G, Sampson NA and Kessler RC. (2006) Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys. *International journal of methods in psychiatric research,* 15, 167-180.

Helzer JE, Robins LN, McEvoy LT, Spitznagel EL, Stoltzman RK, Farmer A and Brockington IF. (1985) A comparison of clinical and diagnostic interview schedule diagnoses. Physician reexamination of lay-interviewed cases in the general population. *Archives of General Psychiatry,* 42, 657-666.

Henningsen P and Lowe B. (2006) Depression, pain, and somatoform disorders. *Current Opinion in Psychiatry,* 19, 19-24.

Hodiamont PPG. (1986) The "what" in psychiatric epidemiology [in Dutch]. In *Searching sick souls [Thesis].* pp. 61-82. KUN-ISG-nr.16, Nijmegen.

Hodiamont PPG, Peer N and Syben N. (1987) Epidemiological aspects of psychiatric disorder in a Dutch health area. *Psychological Medicine,* 17, 495-505.

Hodiamont PPG, Rijnders CATh, Mulder J and Furer JW. (2005) Psychiatric disorders in a Dutch Health Area: a repeated cross-sectional survey. *Journal of Affective Disorders,* 84, 77-83.

Jacobi F, Wittchen HU, Holting C, Hofler M, Pfister H, Muller N and Lieb R. (2004) Prevalence, co-morbidity and correlates of mental disorders in the general population: results from the German Health Interview and Examination Survey (GHS). *Psychological Medicine,* 34, 597-611.

Janca A, Isaac M and Ventouras J. (2006) Towards better understanding and management of somatoform disorders. *International Review of Psychiatry,* 18, 5-12.

Kendell RE. (1993) Diagnosis and classification. In RE Kendell en AK Zealley (Eds.), *Companion to psychiatric studies*. ed. 5 pp. 277-294. Churchill Livingstone, Edinburgh, UK.

Kessler RC, Anthony JC, Blazer DG, Bromet E, Eaton WW, Kendler K, Swartz M, Wittchen HU and Zhao S. (1997) The US National Comorbidity Survey: overview and future directions [editorial]. *Epidemiologia e Psichiatria Sociale,* 6, 4-16.

Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR and Walters EE. (2005a) Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry,* 62, 593-602.

Kessler RC, Chiu WT, Demler O, Merikangas KR and Walters EE. (2005b) Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry,* 62, 617-627.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU and Kendler KS. (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of General Psychiatry,* 51, 8-19.

Maj M. (2005) "Psychiatric comorbidity": an artefact of current diagnostic systems? *British Journal of Psychiatry,* 186, 182-184.

Mayou R, Kirmayer LJ, Simon G, Kroenke K and Sharpe M. (2005) Somatoform disorders: time for a new approach in DSM-V. *American Journal of Psychiatry,* 162, 847-855.

McConnell P, Bebbington P, McClelland R, Gillespie K and Houghton S. (2002) Prevalence of psychiatric disorder and the need for psychiatric care in Northern Ireland. Population study in the District of Derry. *British Journal of Psychiatry,* 181, 214-219.

McLeod JD, Turnbull JE, Kessler RC and Abelson JM. (1990) Sources of discrepancy in the comparison of a lay-administered diagnostic instrument with clinical diagnosis. *Psychiatry Research,* 31, 145-159.

Narrow WE, Rae DS, Robins LN and Regier DA. (2002) Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 surveys' estimates. *Archives of General Psychiatry,* 59, 115-123.

Nienhuis FJ, and Giel R (2006) *Mini-SCAN - Mini-Schedules for Clinical Assessment in Neuropsychiatry - 1.0.* Groningen

Ooms I, Woittiez I and van Gameren E. (2002) *Report on the modelling of Mental Health Care for the benefit of a care estimation model [in Dutch]* (Rep. No. Work document 82). SCP, The Hague.

Pincus HA, Tew JD and First MB. (2004) Psychiatric comorbidity: is more less? *World Psychiatry,* 3, 18-23.

Pull CB, Saunders JB, Mavreas V, Cottler LB, Grant BF, Hasin DS, Blaine J, Mager D and Üstün BT. (1997) Concordance between ICD-10 alcohol and drug use disorder criteria and diagnoses as measured by the AUDADIS-ADR, CIDI and SCAN: results of a cross-national study. *Drug and Alcohol Dependence,* 47, 207-216.

Quintana MI, Gastal FL, Jorge MR, Miranda CT and Andreoli SB. (2007) Validity and limitations of the Brazilian version of the Composite International Diagnostic Interview (CIDI 2.1). *Revista Brasileira de Psiquiatria,* 29, 18-22.

Ravelli A, Bijl RV and van Zessen G. (1998) Comorbiditeit van psychiatrische stoornissen in de Nederlandse bevolking; resultaten van de Netherlands Mental Health Survey and Incidence Study (NEMESIS) [in Dutch]. *Tijdschrift voor Psychiatrie,* 40, 531-543.

Regier DA. (2000) Community diagnosis counts [comment]. *Archives of General Psychiatry,* 57, 223-224.

Rief W, Henningsen P and Hiller W. (2006) Classification of somatoform disorders. *American Journal of Psychiatry,* 163, 746-747.

Rijnders CATh and Furer JW. (2003) *Other people, different care: a study on determinants and consequences of comorbidity in the Nijmegen Health Area [in Dutch]* (Rep. No. 64496). Department of Social Medicine, RU, Nijmegen.

Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA and others. (1988) The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry,* 45, 1069-1077.

Roca M, Gili M, Ferrer V, Bernardo M, Montano JJ, Salva JJ, Flores I and Leal S. (1999) Mental disorders on the island of Formentera: prevalence in general population using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Social Psychiatry and Psychiatric Epidemiology,* 34, 410-415.

Rodgers B and Mann SA. (1986) The reliability and validity of PSE assessments by lay interviewers: a national population survey. *Psychological Medicine,* 16, 689-700.

Rosenman SJ, Korten AE and Levings CT. (1997) Computerised diagnosis in acute psychiatry: validity of CIDI-Auto against routine clinical diagnosis. *Journal of Psychiatric Research,* 31, 581-592.

Sartorius N, Shapiro R, Kimura M and Barrett K. (1972) WHO international pilot study of schizophrenia. *Psychological Medicine,* 2, 422-425.

Shibre T, Kebede D, Alem A, Negash A, Kibreab S, Fekadu A, Fekadu D, Jacobsson L and Kullgren G. (2002) An evaluation of two screening methods to identify cases with schizophrenia and affective disorders in a community survey in rural Ethiopia. *International Journal of Social Psychiatry,* 48, 200-208.

Slade T and Andrews G. (2002) Empirical impact of DSM-IV diagnostic criterion for clinical significance. *The Journal of Nervous and Mental Disease,* 190, 334-337.

Starcevic V. (2006) Somatoform disorders and DSM-V: conceptual and political issues in the debate. *Psychosomatics,* 47, 277-281.

Stein E. (2006) Classification of somatoform disorders. *American Journal of Psychiatry,* 163, 747-748.

Surtees PG, Dean C, Ingham JG, Kreitman NB, Miller PM and Sashidharan SP. (1983) Psychiatric disorder in women from an Edinburgh community: associations with demographic factors. *British Journal of Psychiatry,* 142, 238-246.

Tomov T and Nikolov V. (1990) Reliability of SCAN categories and scores. Results of the field trials. In CN Stefanis (Ed.), *Psychiatry: a world perspective*. pp. 107-117. Elsevier Science Publishers, Amsterdam.

Üstün B, Compton W, Mager D, Babor T, Baiyewu O, Chatterji S, Cottler L, Gogus A, Mavreas V, Peters L, Pull C, Saunders J, Smeets R, Stipec MR, Vrasti R, Hasin D, Room R, Van den Brink W, Regier D, Blaine J, Grant BF and Sartorius N. (1997) WHO Study on the reliability and validity of the alcohol and drug use disorder instruments: overview of methods and results. *Drug and Alcohol Dependence,* 47, 161-169.

van den Akker M, Buntinx F and Knottnerus JA. (1996) Comorbidity or multimodality: what's in a name? A review of literature. *European Journal of General Practice,* 2, 65-70.

van Limbeek J, van den Berg CEA, Sergeant JA, Geerlings PJ and Fransman JML. (1994) *Patient, disorder and care [in Dutch]*. GG&GD and University of Amsterdam, Amsterdam.

van Os J, Driessen G, Gunther N and Delespaul P. (2000) Neighbourhood variation in incidence of schizophrenia. Evidence for person-environment interaction. *British Journal of Psychiatry,* 176, 243-248.

Verhaak PF, Hoeymans N, Garssen AA and Westert GP. (2005) Mental health in the Dutch population and in general practice: 1987-2001. *British Journal of General Practice,* 55, 770-775.

Waller E and Scheidt CE. (2006) Somatoform disorders as disorders of affect regulation: a development perspective. *International Review of Psychiatry,* 18, 13-24.

Wing JK. (1996) SCAN and the PSE tradition. *Social Psychiatry and Psychiatric Epidemiology,* 31, 50-54.

Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D and Sartorius N. (1990) SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry,* 47, 589-593.

Wing JK, Birley JL, Cooper JE, Graham P and Isaacs AD. (1967) Reliability of a procedure for measuring and classifying "present psychiatric state". *British Journal of Psychiatry,* 113, 499-515.

Wing JK, Cooper JE and Sartorius N. (1974) *The desciption and classification of psychiatric symptoms. An instruction manual for the PSE and CATEGO system*. Cambridge University Press, Cambridge.

Wittchen HU. (1994) Reliability and validity studies of the WHO--Composite International Diagnostic Interview (CIDI): a critical review. *Journal of Psychiatric Research,* 28, 57-84.

Wittchen HU, Üstün TB and Kessler RC. (1999) Diagnosing mental disorders in the community. A difference that matters? [editorial]. *Psychological Medicine,* 29, 1021-1027.

11 Appendix

11.1 Psychiatric disorders in a Dutch health area: a repeated cross-sectional survey

P.P.G. HODIAMONT[1,2,3], C.A.Th. RIJNDERS[1,3] J. MULDER[1] and J.W. FURER[1]

[1] Department of Social Medicine, Radboud University, Nijmegen; [2] Department of Clinical Health Psychology, University of Tilburg, The Netherlands; [3] Institute for Mental Health Care, Midden Brabant, Tilburg, The Netherlands

Journal of Affective Disorders (2005), 84, 77-83

ABSTRACT

Background. Decades of psychiatric epidemiological research have shown a wide variation in prevalence rates, but a consistent relationship between psychiatric disorder and sociodemographic variables. In this repeated cross-sectional survey, the prevalence rate of psychiatric disorders and their distribution in the general population of the same area were determined in 1983 and 1997.

Methods. At an interval of 14 years, two two-phase studies of psychiatric prevalence were carried out among the inhabitants of a Dutch Health Area (Nijmegen). In phase 1, the GHQ-30 was administered to a random sample of persons, and in phase 2, the respondents were interviewed by means of a clinical semi-structured interview. Only the phase-1 data will be reported here.

Results. The mean overall GHQ score changed significantly, from 3.1 (± 1.0) in 1983 to 4.6 (± 1.8) in 1997.

With respect to bivariate analyses, higher score rates were found consistently in the age category ≥ 50 years among divorced persons, poorly educated persons, the unemployed or chronically ill and urban residents. With regard to multivariate analyses (second-order effect), however, the variance explained by these socio-demographic variables doubled, revealing the importance of complex interactions.

Limitations. Our objective of ensuring identical designs in 1983 and 1997 could not be completely achieved.

Conclusions. Over the interval in question, psychiatric prevalence had increased in all sociodemographic categories, despite the improved socio-economic conditions in the survey population as a whole. The increasing complexity of life is apparently taking its toll, even among the socially best equipped.

Key words:

Psychiatric epidemiology, repeated cross-sectional survey, time-trend analysis

11.1.1 Introduction

Time and again it has been suggested that the prevalence rate of psychiatric disorders is rising owing to noxious factors that characterise the way of life in modern society. In spite of numerous social changes on a global scale, however, various studies have found psychiatric prevalence rates to be stable (Murphy et al. 1984; Nandi et al. 2000) or even declining (Bebbington et al. 1981; Bebbington et al. 1997). Studies on the question of whether psychiatric prevalence rates are changing (Regier et al. 1998) are fraught with methodological difficulties. While a prospective cohort study has the advantage of consistency as far as the population is concerned, secular population changes cannot be examined adequately in this way since the cohort ages over time. A repeated cross-sectional survey, on the other hand, does offer that opportunity, but with the disadvantage of a relative incomparability of the study populations. Because of these difficulties, studies comparing community psychiatric surveys carried out in the same area at different points in time are few.

To determine any change in psychiatric prevalence rates and to test the hypothesis of a consistent association between demographic characteristics and psychopathology (Dohrenwend & Dohrenwend 1974; Surtees et al. 1983), a repeated cross-sectional survey was carried out in the population of the Nijmegen Health Area (NHA) in 1983 (Hodiamont et al. 1987) and 1997. The General Health Questionnaire (GHQ-30) (Goldberg 1972; Goldberg 1978) and the Present State Examination (PSE) (Wing 1974; Wing 1996), combined in a two-phase design, were used to compare case rates and their distribution between 1983 and 1997. We shall restrict this paper to the results from the first phase, i.e. the data gathered with the GHQ-30.

11.1.2 Methods

The sample

Random samples of 4,500 and 4,517 persons from the non-institutionalised Dutch-speaking population of the NHA aged 18-64 years were enrolled in 1983 and 1997. For the 1983 sample (Hodiamont et al. 1987), the distribution of the 3,245 respondents by gender, age, and marital status corresponded to the population distribution according to the 1982 census data of the Central Bureau of Statistics (CBS). A total of 3,232 persons completed the phase-1 interview.

In order to have access to medical data, we recruited the 1997 sample through general practitioners (GPs). Since virtually every inhabitant of the Netherlands is registered in a general practice, the degree of this registration is equivalent to that of the registry offices (Boerma et al. 1993). The gross sample of 4,517 persons, which corresponded closely to the population distribution by age and gender according to the 1997 census data and was distributed equally over the participating general practices, were sent a letter by their GP in which they were asked whether they would consent to take part in the survey on an informed basis. A phase-1 interview was conducted on 1,617 persons. To establish whether the response showed a selection bias and consequently whether the survey sample was an adequate representation of the general population, data on gender, age, degree of urbanisation and the prescription of psychotropic medication (as a parameter of mental illness) were collected (Table 11.1.1). Since no significant difference in prescription by the GP of psychotropic medication was found between the survey sample and the general population, it was concluded that a selection bias with respect to psychiatric disorders was unlikely.

Females and rural dwellers were overrepresented in both survey samples and the elderly in 1997 only. Adjustments for these differences were made in the statistical analyses. Like the 1983 results, the 1997 results may consequently be considered representative of the general population.

The general interview

The phase-1 interview with the subjects in the survey sample covered, among other things, sociodemographic variables and the GHQ-30 items, which constitute a self-administered test that is well established as a detector of psychiatric disorders in community settings.

Table 11.1.1

Comparison of the 1983 and 1997 survey samples (NHA-1 and NHA-2) with CBS data with respect to degree of urbanisation, gender and age; and representativeness of the 1997 survey sample for psychotropic medication prescription in comparison with GP practices.

		1983			1997		
		Survey sample N = 3,245 %	Population N = 247,141 %	χ^2	Survey sample N = 1,617 %	Population N = 294,782 %	χ^2
Urbanisation	Nijmegen	33.6	39.1	p < .001	29.2	34.3	p < .001
	SC *	66.4	60.9		70.8	65.7	
Gender	Males	48.7	50.5	p = .037	44.3	50.4	p < .001
	Females	51.3	49.5		55.7	49.6	
Age (years)	18-19	5.2	5.9	p < .001	1.5	3.6	p < .001
	20-24	14.7	15.3		5.9	11.1	
	25-29	15.5	14.7		9.2	13.2	
	30-34	14.9	14.0		13.2	13.3	
	35-39	12.7	11.5		15.9	12.5	
	40-44	10.3	9.3		13.0	11.8	
	45-49	6.9	8.5		12.7	11.7	
	50-54	7.6	7.8		11.6	8.7	
	55-59	7.4	7.1		9.2	7.5	
	60-64	4.9	5.9		7.7	6.7	
					Survey sample N = 1,617	All patients N = 80,315	
Prescriptions							
	Antidepressants				3.4	3.4	.98 (NS)
	Benzodiazepines				10.3	9.5	.21 (NS)
	Antipsychotics				0.4	0.8	.07 (NS)
	Analgesics				10.3	9.6	.33 (NS)
	Other psychotropic drugs				2.7	2.7	.88 (NS)

* the small towns and villages around the city of Nijmegen (SC)

Analyses

Version 8 of the SAS package was used for all statistical analyses.

Gender, age, marital status, education, employment status and degree of urbanisation were tested to determine whether they had a significant impact on the mean GHQ scores. For the differences between the survey sample and the general population with respect to gender, degree of urbanisation and age, adjustments were made in the statistical analyses. A MANOVA was performed to identify both main and interaction effects.

11.1.3 Results

Sociodemographic shifts

From the CBS data (Table 11.1.1), it is obvious that the NHA population had increased by almost 20% in the 14-year interval in question, mainly (91% of the growth) on account of the newly built areas in the small towns and villages around the city of Nijmegen, i.e. the surrounding communities (SC). The population also showed marked aging.

Our data (e.g., Table 11.1.2) show that the sociodemographic profile of the population changed markedly after 1983. With reference to marital status, there was an overall increase in the number of divorced persons and a decrease in the number of singles, while, in the rural areas only, there was an increase in the number of persons living with a partner. The mean highest completed level of education increased, which resulted in a smaller contribution by persons who had only a primary school education.

The number of persons in remunerated employment grew substantially and the percentage of housewives in full-time employment was halved. The percentage of chronically ill or unemployed persons decreased for males, but was broadly the same for females.

Table 11.1.2
Sociodemographic distribution in 1983 and 1997 (Males)

		Males							
		Nijmegen				SC			
		1983		1997		1983		1997	
		512		191		1060		519	
		N	%	N	%	N	%	N	%
Marital status	Single	123	24.0	47	24.6	227	21.4	60	11.6
	Living with partner	369	72.1	131	68.6	814	76.8	436	84.0
	Divorced or Widowed	20	3.9	13	6.8	19	1.8	23	4.4
Education	Primary or none	200	39.1	43	22.5	537	50.7	177	34.1
	Intermediate	147	28.7	72	37.7	265	25.0	211	40.7
	Post-second.	165	32.2	76	39.8	258	24.3	131	25.2
Employment status	Employed or Retired	310	60.6	158	82.7	799	75.4	458	88.3
	Chron. ill or Unemployed	110	21.5	16	8.5	189	17.8	45	8.7
	Householder or student	92	18.0	17	8.9	72	6.8	15	2.9

In some subgroups, due to missing data, the cumulative N is lower than the N of the survey sample stated in the heading of the table.

Table 11.1.2 (continued)
Sociodemographic distribution in 1983 and 1997 (Females)

		Females							
		Nijmegen				SC			
		1983		1997		1983		1997	
		571		276		1089		619	
		N	%	N	%	N	%	N	%
Marital status	Single	114	20.0	52	18.9	149	13.7	40	6.5
	Living with partner	408	71.5	199	72.4	880	80.8	544	87.9
	Divorced or Widowed	49	8.6	24	8.7	60	5.5	35	5.7
Education	Primary or none	241	42.2	56	20.3	580	53.3	176	28.5
	Intermediate	196	34.3	70	25.4	346	31.8	337	54.5
	Post-second.	134	23.5	150	54.4	163	15.0	105	17.0
Employment status	Employed or Retired	207	36.3	197	71.6	399	36.6	366	59.1
	Chron. ill or Unemployed	32	5.6	17	6.2	42	3.9	21	3.4
	Householder or student	332	58.1	61	22.2	648	59.5	233	37.6

In some subgroups, due to missing data, the cumulative N is lower than the N of the survey sample stated in the heading of the table.

Prevalence

Table 11.1.3 presents the distributions of the GHQ-30 scores by gender, degree of urbanisation (Nijmegen and SC), and year of survey (NHA-1 - 1983 and NHA-2 - 1997). In both surveys, females scored higher than males and Nijmegen respondents scored higher than those of the SC. The mean scores increased significantly over time ($.01 < p < .001$), most markedly for females in the city of Nijmegen (+2.7). While the differences between Nijmegen and the SC were minimal in 1983, in 1997 the mean score for Nijmegen was clearly and significantly higher than that for the SC ($p = .0017$).

The increase concerned the items on social dysfunction in particular, the scores for items on social interaction having deteriorated markedly. On the other hand, the scores for items on socio-economic prospects were unchanged or even better.

In short, although the aforementioned shift from 1983 to 1997 indicated that the socio-economic status of the population had improved, its health status in terms of the GHQ-scores had worsened.

Table 11.1.3
Frequency distribution of GHQ-30 scores by gender, degree of urbanisation and year of survey*

	Males				Females			
	Nijmegen		SC		Nijmegen		SC	
	1983	1997	1983	1997	1983	1997	1983	1997
	N = 512	N = 191	N = 1060	N = 519	N = 571	N = 276	N = 1089	N = 619
GHQ score	%	%	%	%	%	%	%	%
0 – 2	62.6	54.2	64.7	63.2	59.9	45.2	63.3	55.8
3 – 5	18.4	18.1	17.5	14.8	18.5	18.8	18.1	14.9
6 – 7	5.7	5.0	5.9	6.1	7.0	6.1	6.1	5.0
8 – 9	2.5	3.8	4.4	4.1	4.9	6.1	4.1	6.3
≥ 10	10.8	18.9	7.4	11.8	9.8	23.9	8.4	18.1
Mean score	3.2	5.2	2.9	3.6	3.4	6.1	3.0	4.7
S.D.	2.1	3.8	1.9	2.9	2.1	4.0	2.0	3.5

* Adjusted for age

The relation between GHQ data and sociodemographic variables (see Table 11.1.4)

Gender

Whereas the mean GHQ score did not differ for gender in 1983, the increase in mean score in 1997 was twice as high for females as for males, resulting in a significant difference.

Age

Our conclusion in 1983 that 'the case prevalence more or less increases with age' should be readjusted. In 1997 the mean GHQ scores were more or less the same in all age categories.

Marital status

The finding in 1983 of a higher mean GHQ score for divorced persons than for those living with a spouse or partner was confirmed in 1997. The difference in the mean GHQ scores between the two years was significantly higher for those living alone or with a spouse or partner than for divorced subjects.

Education

There was a strong inverse relationship between disorder and education. For the higher educated however, not only was there an increase in mean GHQ scores, but this increase was greater than for any of the other educational levels.

Table 11.1.4
Sociodemographic distribution of GHQ scores

	1983				1997			
		Main effects only	**Main + 1st-order [a]**	**Main + 2nd-order [b]**		**Main effects only**	**Main + 1st-order [c]**	**Main + 2nd-order [d]**
Proportion explained		7.0%	9.0%	10.5%		6.0%	10.0%	13.1%
variance	Mean GHQ	p	p	p	Mean GHQ	p	p	p
Gender								
Male	3.0				4.0			
Female	3.1	--	--	--	5.1	.0017	--	--
Age								
18-34	2.8				4.7			
35-49	3.1				4.5			
50-64	3.9	.0853	--	--	4.8	--	.0456	.0258
Marital status								
Single	2.8				5.0			
Living with a spouse/partner	3.0				4.5			
Divorced or Widowed	5.5	<.0001	<.0001	<.0001	5.9	--	.0142	.0313
Education								
Primary or none	3.7				5.5			
Intermediate	2.9				4.4			
Post-second.	2.1	<.0001	.0068	.0038	4.1	.0045	.0017	.0066
Employment status								
Employed or Retired	2.4				4.1			
Chron. ill or Unemployed	5.4				8.7			
Householder or student	3.3	<.0001	<.0001	<.0001	5.3	<.0001	<.0001	<.0001
Degree of urbanisation								
Nijmegen	3.4				5.7			
SC	3.0	.0334	--	--	4.2	<.0001	<.0001	.0002

[a] Employment status * Gender .0055; Employment status * Age .0156; Employment status * Education .0008

[b] Gender * Education .0029; Employment status * Marital status .0393; Employment status * Gender .0071; Employment status * Education .0086; Education * Marital status .0387; Gender * Education *Marital status .0016; Employment status * Age * Degree of urbanisation .0150; Age * degree of urbanisation * Education .0388

[c] Gender * Education .0232; Employment status * Degree of urbanisation .0018; Employment status * Marital status .0219

[d] Employment status * Degree of urbanisation .0001; Employment status * Marital status .0389; Gender * Education * Employment status .0043; Education * Marital status* Degree of urbanisation .0496.

Employment status

A significant difference in rates of disorder was found in the comparison of the unemployed or chronically ill with the other categories, irrespective of gender, degree of urbanisation and year of inquiry ($p < .05$).

Degree of urbanisation

The difference in the mean GHQ scores between the city of Nijmegen and the SC almost quadrupled from 1983 to 1997. With respect to the distribution of psychiatric disorder, degree of urbanisation, consequently, came to the fore in 1997.

11.1.4 Discussion

The objectives of this paper were to compare the prevalence rate of the GHQ scores and their distributions between 1983 and 1997. At first sight, the notable non-response in 1997 appears to contradict the assumption of a representative survey sample and limit the comparability of the samples between 1983 and 1997. Potential respondents seemed deterred by the exhaustive information campaign deemed necessary for informed consent by the medical ethics committee. On the other hand, our sampling design, in which GP practices were used, enabled us to check for selectivity by means of study-relevant variables, and there was no evidence of a selection bias. Consequently, we assume that in 1997 too, the sample was a fair representation of the general population.

Using a cut–off point of 5/6, we estimated the prevalence rate of the GHQ caseness at 19.0% in 1983 and 28.9% in 1997 in representative samples of a Dutch general population, aged 18-64 years. Thus, we found the prevalence rate to have risen by about 50% for the health area as a whole. With respect to bivariate analyses, higher score rates were consistently found in the age category ≥ 50 years among divorced persons, the poorly educated, the unemployed or chronically ill, and urban residents. In the multivariate analysis on a second level, however, the variance explained by these sociodemographic variables doubled, revealing the importance of complex interactions. The mean GHQ score for females in Nijmegen, for instance, almost doubled from 1983 to 1997.

All factors considered, we may conclude that while the distribution of psychiatric disorder did not change, its prevalence rate grew substantially in the NHA from 1983 to 1997. This increase in prevalence rate can be explained in terms of social change - which was definitely substantial in the Nijmegen Health Area, and especially in the city. Indeed, Murphy et al. (1984) hypothesised that such change is potentially noxious for mental health. Over time, the rates increased in all sociodemographic categories. Despite the improved socio-economic conditions in the survey population as a whole, the increasing complexity of modern life is apparently taking its toll, even among those best equipped to deal with it, in particular the highest educated and married persons.

References

Bebbington PE, Hurry J, Tennant C, Sturt E and Wing JK. (1981) Epidemiology of mental disorders in Camberwell. *Psychological Medicine,* 11, 561-579.

Bebbington PE, Marsden L and Brewin CR. (1997) The need for psychiatric treatment in the general population: the Camberwell Needs for Care Survey. *Psychological Medicine,* 27, 821-834.

Boerma WGW, de Jong FAJM and Mulder PH. (1993) The Netherlands. In *Health Care and General Practice across Europe*. pp. 77-80. Nivel (Netherlands Institute of Primary Health Care), Utrecht.

Dohrenwend BP and Dohrenwend BS. (1974) Social and cultural influences on psychopathology. *Annual Review of Psychology,* 25, 417-452.

Goldberg DP. (1972) *The detection of psychiatric illness by questionnaire*. Oxford University Press, London.

Goldberg DP. (1978) *Manual of the General Health Questionnaire*. NFER Publishing Company, Windsor.

Hodiamont PPG, Peer N and Syben N. (1987) Epidemiological aspects of psychiatric disorder in a Dutch health area. *Psychological Medicine,* 17, 495-505.

Murphy JM, Sobol AM, Neff RK, Olivier DC and Leighton AH. (1984) Stability of prevalence. *Archives of General Psychiatry,* 41, 990-997.

Nandi DN, Banerjee G, Mukherjee SP, Ghosg A, Nandi PS and Nandi S. (2000) Psychiatric morbidity of a rural Indian community. *British Journal of Psychiatry,* 176, 351-356.

Regier DA, Kaelber CT, Rae DS, Farmer ME, Knauper B, Kessler RC and Norquist GS. (1998) Limitations of diagnostic criteria and assessment instruments for mental disorders. Implications for research and policy. *Archives of General Psychiatry,* 55, 109-115.

Surtees PG, Dean C, Ingham JG, Kreitman NB, Miller PM and Sashidharan SP. (1983) Psychiatric disorder in women from an Edinburgh community: associations with demographic factors. *British Journal of Psychiatry,* 142, 238-246.

Wing JK. (1974) *Measurement and classification of psychiatric symptoms*. Cambridge University Press, Cambridge.

Wing JK. (1996) SCAN and the PSE tradition. *Social Psychiatry and Psychiatric Epidemiology,* 31, 50-54.

11.2 Conversion table SCAN/PSE-10 to PSE-9

PSE-9	**SCAN (PSE-10)**
1 Physical fitness (subjective) 0 1 2 3	2.001 1 2 3 4
2 Physical illness or disabilities (obj.) 0 1 2 [xxxii] 3	2.004 0 1 2 -
3 Psychosomatic problems 0,1	 Not convertible
4 Worrying 0 1 2	3.001 0,1 2 3
5 Tension pains 0 1 2	3.005 0,1 2 3
6 Fatigability and exhaust 0 1 2	3.007 0,1 2 3
7 General muscular tension 0 1 2	3.003 0,1 2 3
8 Restlessness 0 1 2	3.006 0,1 2 3
9 Hypochondriacal preoccupation 0 1 2	2.084 0 1 2
10 Nervous tension 0 1 2	3.002 0,1 2 3
10-a Sensitivity to noise 0 1 2	3.008 0,1 2 3

PSE-9	SCAN (PSE-10)
11Free-floating anxiety 0 1 2	4.023 0,1 2 3
12 Anxious foreboding 0 1 2	4.024 0,1 2 3
13 Delusional anxiety 0,1,2	Not convertible
14 Panic attacks 0 1 2	4.020 00 >00 ≤04 >04
15 Situation related anxiety 0 1 2	4.027/028/029/030/031/032 0 - Highest score counts 1,2
16 Social anxiety 0 1 2	4.033/034/035/036 0 - Highest score counts 1,2
17 Specific phobia 0 1 2	4.037/038/039/040/041/042/043/044/045 0 - Highest score counts 1,2
18 Avoidance 0 1 2	4.047/050/053 0,1 2 Highest score counts 3
19 Subjective inefficient thinking 0 1 2	7.003 0,1 2 3
20 Loss of concentration 0 1 2	7.002 0,1 2 3
21 Interference due to worrying 0 1 2	3.015 0,1 2 3
22 Loss of interests 0 1 2	7.004 0,1 2 3

PSE-9	SCAN (PSE-10)		
23 Depressed mood	6.001		
0	0,1		
1	2		
2	3		
24 Loss of hope for the future	6.006		
0	0,1		
1	2		
2	3		
25 Suicide	6.011		
0	0,3		
1	1		
2	2		
3	4		
26 Depression or anxiety primary	6.022		
0	1		
1	2		
2	3		
27 Morning depression	6.009		
0	0		
1	-		
2	1		
28 Social withdrawal	6.016		
0	0,1		
1	2		
2	3		
29 Loss of self-esteem	6.017		
0	0,1		
1	2		
2	3		
30 Loss of self-confidence	6.015		
0	0,1		
1	2		
2	3		
31 Simple ideas of reference	3.010		
0	0,1		
1	2		
2	3		
32 Guilty ideas of reference	6.014		
0	0,1		
1	2		
2	3		
33 Pathological guilt	6.013		
0	0,1		
1	2		
2	3		
34 Loss of weight	8.005		8.006
0	0		0
1	1,2,3	*AND*	1
2	1,2,3		2,3,4

PSE-9	SCAN (PSE-10)
35 Delayed sleep 0 1 2	8.011 0 1 2
36 Subjective feeling of retardation / loss of drive 0 1 2	7.005/006 0,1 2 Highest score counts 3
37 Early waking 0 1 2	8.014 0 1 2,3
38 Loss of libido 0 1 2	8.024 0,1 2 3
39 Premenstrual aggravation 0,1,2	 Not convertible
40 Irritability 0 1 2 3	3.009 0 1 2 3
41 Expansive / elevated mood 0 1 2	10.001 0 1,2 3
42 Pressing and racing thoughts 0 1 2	10.004 0,1 2 3
43 Grandiosity 0 1 2	10.010 10.012 0,1 0,1,2 2 - Highest score counts 3 3
44 Obsessional checking and repeating 0 1 2	5.002 0 (score 1 not in use) 2 3
45 Obsessional cleanliness 0 1 2	5.005 0 (score 1 not in use) 2 3
46 Obsessional thoughts/ruminations 0 1 2	5.004/006 0 (score 1 not in use) 2 3

PSE-9	SCAN (PSE-10)		
47 Derealisation	16.006/007		3.012[xxxiii]
0	0		0
1	1	Highest score counts	1
2	2,3		2,3
48 Depersonalisation	16.008/009/010		3.012[xxxiv]
0	0		0
1	1	Highest score counts	1
2	2,3		2,3
49 Delusional mood	18.001		
0	0		
1	1		
2	2,3		
50 Heightened perception	16.004		
0	0		
1	-		
2	1,2,3		
51 Dulled perception	16.003		
0	0		
1	-		
2	1,2,3		
52 Changing perceptions	16.002		
0	0		
1	-		
2	1,2,3		
53 Changed perception of time	16.005		
0	0		
1	-		
2	1,2,3		
54 Feeling of loss of feeling	6.007		
0	0,1		
1	2		
2	3		
55 Thought insertion	18.006		
0	0		
1	-		
2	1,2,3		
56 Loud thought/ though broadcast	18.004	18.007	
0	0	0	
1	1,2,3	-	Highest score counts
2	-	1,2,3	
57 Thought echo / commentary	18.005	18.008	
0	0	0	
1	1,2,3	-	Highest score counts
2	-	1,2,3	
58 Thought block/withdrawal	18.009	18.010	
0	0	0	
1	1,2,3	-	Highest score counts
2	-	1,2,3	

PSE-9	SCAN (PSE-10)
59 Delusion of thoughts being read 0 1 2	18.003 0 - 1,2,3
60 Non-verbal auditory hallucinations 0 1 2	17.003 0 - 1,2,3
61 Mood congruence of auditory hallucinations 0 1 2 3	17.011 / 17.010 0,2 / 0,2,3 - / 1 1 / - Highest score counts 3 / -
62 Auditory hallucinations, 3rd person 0 1 2 [xxxv]	17.008 / 17.009 0,1 / 0,1,2,3 - / - 2,3 AND 4,5
63 Auditory hallucinations, 2rd person 0 1 2 [xxxvi]	17.009 / 17.010 0,3,4,5 / 0,1,2,3 - / - 1,2 *AND* 4,5,9
64 Dissociative hallucinations 0 1 2	17.020 / 20.084 0 / - 1,2,3 / 2,3 Highest score counts 1,2,3 / 0,1
65 Internal hallucinations 0 1 2 3	17.007 0 2 3 1
66 Visual hallucinations 0 1 2	17.015 / 17.016 / 17.017 0 / 0 / 0 Highest 1,2,3 / - / - score - / 1,2,3 / 1,2,3 counts
67 Visual hallucinations during delirium 0 1 2	ALS 66 / 21.119/121/123 [xxxvii] 0 / 0,1 1 *AND* 2 2 / 2
68 Olfactory hallucinations, delusions associated with smell 0 1 2	17.022 / 17.023 0 / 0 1,2,3 / - Highest score counts - / 1,2,3
69 Delusions of smelling 0 1 2	17.024 / 17.025 0 / 0 1,2,3 / - Highest score counts - / 1,2,3

PSE-9	SCAN (PSE-10)	
70 Other hallucinations and delusions	17.026/028/030	17.027/029
0	0	0
1	1,2,3	- Highest score counts
2	-	1,2,3
71 Replacement of will by external force	18.012	
0	0	
1	-	
2	1,2,3	
72 Delusions of reference	19.004	
0	0	
1	-	
2	1,2,3	
73 Delusional misinterpretation	19.005	
0	0	
1	-	
2	1,2,3	
74 Delusions of persecution	19.012	
0	0	
1	-	
2	1,2,3	
75 Delusions concerning special help		
0, 1, 2	Not convertible	
76 Delusions of grandiose abilities	10.016	19.029
0	0	0
1	-	- Highest score counts
2	1,2,3	1,2,3
77 Delusions of grandiose identity	10.017	19.030
0	0	0
1	-	- Highest score counts
2	1,2,3	1,2,3
78 Religious delusions	19.021	
0	0	
1	-	
2	1	
79 Delusional paranormal explanations	19.022	
0	0	
1	-	
2	1	
80 Delusional physical explanations	19.023	
0	0	
1	-	
2	1	
81 Delusions of external forces	18.012	19.019
0	0	0
1	-	- Highest score counts
2	1,2,3	1,2,3

PSE-9	SCAN (PSE-10)		
82 Delusional perception	19.009		
0	0		
1	-		
2	1,2,3		
83 Sub cultural defined delusions	19.024		
0	0		
1	-		
2	-		
3	1		
84 Delusional jealousy	19.014		
0	0		
1	-		
2	1,2,3		
85 Delusions of pregnancy	19.016		
0	0		
1	-		
2	1,2,3		
86 Delusions concerning sexuality and love	17.027	19.017/018	
0	0	0	
1	-	-	Highest score counts
2	1,2,3	1,2,3	
87 Fantastic delusions	19.019		
0	0		
1	-		
2	1,2,3		
88 Delusions of guilt	19.025	6.018	
0	0	0	
1	-	-	Highest score counts
2	1,2,3	1,2,3	
89 Delusions concerning appearance	19.031	16.012	
0	0	0	
1	-	-	Highest score counts
2	1,2,3	1,2,3	
90 Delusions of depersonalisation or nihilism	16.013	19.032	
0	0	0	
1	-	-	Highest score counts
2	1,2,3	1,2,3	
91 Hypochondriacal delusions	19.027/028		
0	0		
1	-		Highest score counts
2	1,2,3		
92 Delusions of catastrophe	19.026		
0	0		
1	-		
2	1,2,3		

PSE-9	SCAN (PSE-10)
93 Systematisation of delusions 0 1 2 3	19.035 0 1 2 3,4
94 Denying and hiding delusions 0, 1, 2, 3	 Not convertible
95 Preoccupation with delusions / hallucinations 0 1 2 → Missing hallucinations 3 4 5	19.036 17.001/19.001 19.020 17.014/21 0 0 - - - 8 - - - - 1,2,3 - 1 - Highest - 1 - - score - 2 2 - counts - 3
96 Actions based on delusions 0 1 2 3	19.040 0 - 1 2,3
97 Fugues, blackout, amnesia 0 1 2 3 → Not convertible	2.104 21.076 0 0 - 1 1 2 - -
98 Drug abuse 0 1 2 Highest score counts 3 4	12.001 12.007 12.008 12.010/011 12.006/009 1,2 - - - - - 1-7 - - - - - 1-7 - - - - - 1-7 - - - - - 1-7
99 Alcohol abuse 0 1 2	11.007 11.010/013/025/031/ 0 - 1,2 - Highest score counts - 1,2,3(,4)
100 Disturbance of consciousness 0 1 2	2.102/104/105/106/107 22.006 15.006 0 - - 1 Highest score counts - - - 1,2 1,2
101 Symptoms of conversion 0 1 2	2.105/109/111/113 22.006 0 - 1 Highest score counts - - 1,2
102 Veiled consciousness during examination 0 1 2	15.006 22.006 0 0 1,2 - Highest score counts - 1,2

PSE-9	SCAN (PSE-10)
103 Organic cognitive impairment (MMSE) 0 1 2 3	21.027 ≥ 23 ≥ 21 EN ≤ 22 ≥ 18 EN ≤ 20 ≤ 17
104 Insight into psychotic symptoms 0 1 2 3 9	24.044 - 1 2 3 0
105 Insight into neurotic symptoms 0, 1, 2, 3	Not convertible
106 Social interference due to neurotic symptoms 0 1 2 3	2.123 3.015 4.048/051/054/060 5.015 6.027/062/072 7.009 8.021/028 9.005/026 10.028/055 11.021/037 12.0361/2/3 2.0491/2/3 0 1 2 Highest score counts 3
107 Social interference due to psychotic symptoms 0 1 2 3	20.048 0 1 2 3
108 Self neglect 0 1 2	22.046 0 1 2
109 Odd appearance 0 1 2	22.041 0 1 2
110 Slowness / inactivity 0 1 2	22.002/003/004/005 0 1 Highest score counts 2
111 Agitation 0 1 2	22.016 0 1 2
112 Over activity/destructiveness 0 1 2	22.017 22.018 0 0 1 - Highest score counts 2 1,2

PSE-9	SCAN (PSE-10)
113 Irreverent behaviour 0 1 2	22.044 0 1 2
114 Distractibility 0 1 2	22.014 0 1 2
115 Embarrassing behaviour 0 1 2	22.042 0 1 2
116 Mannerisms 0 1 2	22.033 0 1 2
117 Stereotypies and tics 0 1 2	22.035 0 1 2
118 Apparently hallucinating behaviour 0 1 2	22.054 0 1 2
119 Catatonic behaviour 0 1 2	22.024/025/027/028/032 24.027 0 0 1 Highest score counts 1 2 2
120 Observed anxiety 0 1 2	23.018 0 1 2
121 Observed depression 0 1 2	23.001 0 1 2
122 Histrionic behaviour 0 1 2	22.045 0 1 2
123 Observed elated mood 0 1 2	23.006 0 1 2
124 Observed irritable mood 0 1 2	23.007 0 1 2

PSE-9	SCAN (PSE-10)
125 Suspiciousness 0 1 2	23.019 0 1 2
126 Perplexity 0 1 2	23.020 0 1 2
127 Lability of mood 0 1 2	23.021 0 1 2
128 Blunting 0 1 2	23.012 0 1 2
129 Incongruity off affect 0 1 2	23.013 0 1 2
130 Slow speech 0 1 2	22.007 0 1 2
131 Pressure of speech 0 1 2	24.007 0 1 2
132 Non-social speech 0 1 2	24.032 0 1 2
133 Muteness 0 1 2	24.027 0 1 2
134 Restricted quantity of speech 0 1 2	24.026 0 1 2
135 Neologisms, idiosyncratic use of words 0 1 2	24.021 0 1 2
136 Incoherence of speech 0 1 2	24.022 0 1 2

PSE-9	SCAN (PSE-10)		
137 Flight of ideas	24.008		
0	0		
1	1		
2	2		
138 Poverty of content of speech	24.025		
0	0		
1	1		
2	2		
139 Misleading answers	24.031		
0	0		
1	1		
2	2		
140 Adequacy of interview	13.125	20.114	
0	0	0	
1	1	1	Highest score counts
2	2	2	
3	3	3	

December 5, 1997
C.A.Th Rijnders
F.J. Nienhuis

Summary

The diagnosis of psychiatric disorders rests on the clinical assessment or professional objectification mainly of subjective experiences of the patient. This clinical judgement implies a professional answer to the question whether the presenting signs and symptoms (still) can be regarded as part of normal life or should be assessed as pathological. Clinical judgement is all the more important since psychiatric disorders themselves interfere with the patient's ability to estimate this difference correctly.

Psychiatric epidemiology turns around the question of what condition occurs in whom, where and when. If it is to be established properly whether a patient has a particular condition, the whole complex of case definition, case identification and case finding has to be taken into account, in which event case definition has to be valid and unambiguous, case identification based on clinical relevancy and the process of case finding reliable and valid. Prevalence rates are often not as important or clinically relevant as they seem to be at first glance. Although questionnaires and fully structured interview instruments are frequently used for case detection in general population surveys, questions can be raised about their clinical validity.

The chapters of this thesis are grouped into four parts. Part I consists of four chapters of a theoretical nature and culminates in a description of the research objectives, Part II of two empirical chapters on psychiatric epidemiology and Part III of three chapters on the psychometric properties and aspects of the clinical judgement of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). Finally, in Part IV the results are discussed, the research questions answered and recommendations made.

Part I

In chapter 1, the history of the Nijmegen Health Area Project (NHA) is described, followed by a brief outline of the thesis.

In chapter 2, on psychiatric epidemiology, first the definition and history of this discipline are given, and then the importance is explored of causality and the operational aspects of case identification with reference to the differences between the clinical/objective SCAN/PSE[1] tradition and the respondent-based/subjective CIDI-DIS[2] tradition.

In chapter 3, which addresses diagnostics in psychiatry, the three concepts of disease, illness and sickness are introduced and brought into relation with clinical judgement and classification. The latter converge in the diagnostic process, which in psychiatry is hindered for most disorders by a lack of infallible

[1] Present State Examination
[2] Composite International Diagnostic Interview-Diagnostic Interview Schedule

criteria. This fact underlines the importance of the psychometric quality of the diagnostic instruments used: these should assess accurately the clinically relevant signs and symptoms. The psychometric evaluation of the diagnostic process is explained and an account is given of the choice of the SCAN, a clinically oriented semi-structured diagnostic interview that yields DSM-IV diagnoses.
In chapter 4, the research objectives are set out in terms of research questions. The thesis uses descriptive epidemiological data generated by the SCAN-2.1 through focusing on the relevance of clinical judgement in psychiatric diagnoses, the objective being to improve the quality of the diagnostic process and consequently of the quality of epidemiological data as well.

Part II

In chapter 5 the monitoring study of the NHA project (a repeated cross-sectional study) is described. The objectives were to compare the case rates and their distribution in 1983 and 1997, to explore the impact of urbanisation on these findings, and to investigate possible time trends in terms of cohort, period, and age effects on psychiatric prevalence. Through logistic regression, the one-month prevalence rate of neurotic and functional psychotic caseness was estimated at 7.8 $\pm$ 2.3% (after reanalysis) in 1983 and at 11.9 $\pm$ 2.7% in 1997 in representative samples of the Dutch general population aged 18-64 years. It is thus seen that the prevalence rate of psychiatric disorders had risen by about 50% for the NHA as a whole. For both years, a more or less consistent relationship was found between psychiatric disorders and sociodemographic variables, with a higher prevalence rate for underprivileged groups (the unemployed, chronically ill and poorly educated). The apparent difference in the influence of degree of urbanisation on the rate of psychiatric disorders over the 14-year interval (from 1983 to 1997) tended to decline. No age nor cohort effects were found, but the increase in prevalence rate from 1983 to 1997 may have been caused by several period effects. These concerned both the population as a whole, because of overall changes in social roles and a modified health policy, and in addition the rural communities, owing to the one-way migration from the city and possibly also the imminent inundation of the homes and property of the rural dwellers.
In chapter 6 the mapping study of the NHA-2 is presented. The objectives of this study were: to assess, by means of the SCAN, the 1997 prevalence rates of psychiatric disorders, their sociodemographic distribution and the comorbidity rate in the general population; to study the influence of three newly determined diagnostic categories (sleep, dissociative and somatoform disorders) on prevalence, distribution and comorbidity; and to compare the NHA-2 data with data from the NEMESIS study. The 1-month overall prevalence rate of

psychiatric morbidity in the NHA was 17.6% (± 2.6%). When the three newly determined diagnostic categories were left out, the prevalence rate decreased by one third to 11.9% (±2.5%). With respect to the sociodemographic distribution of psychiatric caseness, the prevalence rate of caseness was significantly higher for females, for subjects living in the city of Nijmegen, the age group 45-54 years, divorced, unemployed and chronically ill subjects and those in the lower education bracket as compared with the subjects in other sociodemographic subgroups. When the three newly determined categories were omitted, the sociodemographic distribution of caseness was characterised by about the same pattern. With respect to comorbidity, 26.1% of the total number of cases had more than one diagnosis. The comorbidity rate for the three newly determined categories themselves was 36.7%. When the survey sample (with respect to age) and diagnostic range in the NHA-2 were reduced to the extent adopted in the NEMESIS study, the prevalence rate for the NHA-2 was 11.7% (±2.2%), which was substantially lower than the 16.5% (±0.4%) prevalence rate found in the NEMESIS study.
The use of the SCAN-2.1 and the integral range of the DSM-IV resulted in a prevalence rate to which the diagnostic categories newly determined in the general population contributed substantially. The sociodemographic distribution of psychiatric disorders turned out to be as Dohrenwend and Dohrenwend had described it years ago. Despite the use of the complete diagnostic range, comorbidity rates in the NHA-2 can be characterised as moderate, notwithstanding the boosting effect of the three newly determined diagnostic categories. Fully structured interviews like the ones used in the NEMESIS study yield higher caseness rates than clinical interviews like the SCAN-2.1.

Part III

In chapter 7 the psychometric properties of the SCAN are described. This study is the first to test the psychometric properties of the SCAN-2.1 for most of the disorders covered and was carried out prior to the NHA-2 study. Two designs were used. In one of these, pairs of independent live interviews with the same respondent were compared (test-retest situation). In the other, ten videotaped interviews conducted by experts were rated by each of the 27 trained interviewers (standardised situation) and the outcomes were compared with those for the other interviewers and also with a reference score. In the test-retest situation the kappa coefficient for diagnostic caseness was substantial (0.62) and for diagnostic categories and diagnostic groups it was moderate to good (0.24 to 0.64). In the standardised situation, in which videotaped interviews by experts were used, both sensitivity and specificity were substantial to almost perfect. The agreement per interviewer with regard to the reference diagnoses ranged from 87% (diagnostic group) to 94% (diagnostic caseness).

Agreement on the syndrome level (without the duration and interference criteria from the DSM-IV) was excellent. Although the instrument is traditionally used by experienced clinicians, this study showed that less experienced (but well trained) interviewers can administer the SCAN reliably.

In chapter 8, the effect of the operationalisation of clinical judgement in the SCAN is elaborated on. As clinical judgement in the SCAN is incorporated into pre-defined rating scales, we were able to process the subclinical scores as if they were equal to the clinical scores, which resulted in the inclusion of all perceived morbidity. When clinical judgement was omitted, the number of positive cases increased by more than one fifth, from 197 to 245 subjects, the most pronounced differences having emerged for somatoform, psychotic and mood disorders. The prevalence rate of caseness without clinical judgement was 21.4%, which is 3.8% higher than when the formal clinical SCAN algorithm was used. The same proportional increase was found for all sociodemographic groups. When clinical judgement was used, one quarter of all subjects diagnosed as a case were comorbid, whereas this was true for one third of all subjects when clinical judgement was left out. In short, clinical judgement interfered significantly with the prevalence rates found, a more pronounced effect having been observed on diagnoses based on subjective experiences than on those rooted in behaviour. When clinical judgement was omitted, the specific diagnoses of lower hierarchical ranking of those subjects who were already a case tended to shift towards the major diagnostic categories. Some specific diagnoses (like somatoform disorder NOS) were extremely sensitive to clinical judgement.

In chapter 9, given that the two instruments probably yield different information, the CIDI is compared with the SCAN in the general population with a view to achieving a better understanding of how each of them works. A total of 139 subjects from the general population were interviewed with the CIDI and the SCAN (in a counterbalanced design). The fully structured CIDI was administered by well-trained lay interviewers and the semi-structured SCAN by well-trained clinically experienced interviewers. Agreement between the CIDI and the SCAN in DSM-IV caseness and diagnosis identification was only moderate. For major affective disorders, agreement with respect to individual signs and symptoms was only poor to fair, whereas canonical correlation analysis disclosed substantial to almost perfect concordance.

Although latent variables identified in the two instruments were closely related, the instruments differed markedly with respect to the identification of cases, diagnoses and signs and symptoms. The goal of making the same DSM-IV diagnoses with the two instruments is compromised by divergent operationalisations of the diagnostic criteria and divergent administration procedures.

Part IV

In chapter 10 the empirical findings reported in Part II (descriptive epidemiology) and Part III (diagnostics: case identification) are viewed from the perspective of the theoretical principles of psychiatric epidemiology and diagnostics (Part I). This discussion culminates in recommendations on how to improve the diagnostic process in psychiatric epidemiology. With respect to descriptive epidemiology, many of the sources of variance mentioned in the literature still play an important role and complicate simple comparisons of gross prevalence rates. In particular, classification systems and interview strategies are important as they tend to recede to the background and become implicit as soon as prevalence rates are calculated and reported. With respect to diagnostics, we conclude that the semi-structured SCAN-2.1 interview, with its clinical judgement, is suitable for case identification in the general population and that the use of clinical judgement has important effects on case identification. Therefore, psychiatric epidemiological population studies should be based on clinically relevant variables, so that clinical judgement is an indispensable instrument for harvesting valid information on diseases.

To conclude, we strongly recommend that both the classification system and the interview strategy be used as an *explicit* frame of reference when it comes to the valuation of research results and that the conclusions from these results should be put into clinical perspective.

Appendix

The appendix is comprised of two parts. The first consists of the published repeated cross-sectional NHA study, based on the General Health Questionnaire (GHQ) data, which was part of the NHA project and used as an important reference in chapter 5. The second presents the conversion table of the SCAN/PSE-10 and PSE-9 scores.

Samenvatting

De diagnostiek van psychiatrische stoornissen stoelt op de klinische inschatting of professionele objectivering van met name subjectieve belevingen bij de patiënt. Dit klinische oordeel impliceert een professioneel antwoord op de vraag of de klachten en verschijnselen als (nog) normaal dan wel als pathologisch moeten worden beschouwd c.q. om welk soort pathologie het dan wel gaat. Omdat psychiatrische stoornissen interfereren met het onderscheidingsvermogen van de patiënt, is dit klinische oordeel des te meer van belang.
De psychiatrische epidemiologie, die berust op de vaststelling wat (welke stoornis) voorkomt bij wie, waar en wanneer, staat of valt met de correcte, dat wil zeggen betrouwbare en valide, bepaling van psychiatrische gevallen. Daarbij speelt het hele diagnostische complex van gevalsdefinitie, gevalsidentificatie en gevalsopsporing een rol: de definitie moet valide en eenduidig zijn, de identificatie gebaseerd op klinische relevantie en het opsporingsproces van gevallen volgens een erkende opzet verlopen. Bij de klinische validiteit van de voor diagnostiek in de open populatie gangbare vragenlijsten en volledig gestructureerde interviews kunnen vraagtekens worden gesteld. De zo gemeten prevalentie is dan ook niet zo eenduidig en klinisch relevant als wel wordt gesuggereerd.
De hoofdstukken van dit proefschrift zijn gegroepeerd in vier delen. Deel I omvat vier hoofdstukken van theoretische aard en culmineert in een beschrijving van de onderzoeksdoelen, deel II twee empirische hoofdstukken over psychiatrische epidemiologie en deel III drie hoofdstukken over de psychometrische en klinische eigenschappen van de Schedules for Clinical Assessment in Neuropsychiatry (SCAN). In deel IV worden tenslotte de resultaten besproken, de onderzoeksvragen beantwoord en aanbevelingen gedaan.

Deel I
In hoofdstuk 1 wordt de geschiedenis van het Regioproject Nijmegen (RN) geschetst en een kort overzicht gegeven van de inhoud van dit proefschrift.
Hoofdstuk 2 gaat over de psychiatrische epidemiologie. Aan de orde komen haar definitie en historie, causaliteits- en operationele aspecten, c.q. het verschil tussen de klinisch/objectieve SCAN/PSE traditie en de respondentgeoriënteerde/subjectieve CIDI/DIS traditie.
In hoofdstuk 3, diagnostiek in de psychiatrie, worden de concepten “disease”, “illness” en “sickness” geïntroduceerd en in verband gebracht met klinisch oordelen en classificeren. De kwaliteit van het diagnostische proces in de psychiatrie is afhankelijk van de deugdelijkheid waarmee klinisch relevante symptomen bepaald kunnen worden, met andere woorden van de

psychometrische kwaliteit van de gebruikte diagnostische instrumenten. In dat kader wordt de keuze verantwoord voor de SCAN, een klinisch georiënteerd, semi-gestructureerd diagnostisch interview dat DSM-IV diagnoses genereert.
In hoofdstuk 4 staan de onderzoeksdoelen beschreven in termen van onderzoeksvragen. Epidemiologische data verzameld met de SCAN-2.1 worden gebruikt om de relevantie van het klinisch oordeel in de psychiatrische diagnostiek te onderzoeken, met als uiteindelijk doel verbetering van de kwaliteit van het diagnostische proces en daarmee van de epidemiologische resultaten.

Deel II

In hoofdstuk 5 over de 'monitoring' studie worden de resultaten van de twee fasen van het RN-project (een herhaald cross-sectioneel onderzoek: RN1 en RN2) beschreven. Dit onderzoek richtte zich op de vergelijking van de prevalentie en de spreiding van psychiatrische stoornissen in 1983 en 1997, op het effect van de urbanisatie op deze bevindingen en op eventuele veranderingen van deze gegevens in termen van cohort-, periode- en leeftijdseffecten. Met behulp van logistische regressie kon de 1-maands prevalentie van neurotische en functionele psychotische stoornissen in een goede afspiegeling van de regionale Nijmeegse open populatie tussen de 18 en 64 jaar (na her-analyse) worden geschat op 7.8 ± 2.3% in 1983 en op 11.9 ± 2.7% in 1997. De prevalentie van psychiatrische stoornissen bleek met ongeveer 50% gestegen voor de Nijmeegse gezondheidsregio als geheel. De relatie tussen de psychiatrische stoornissen en sociodemografische variabelen bleef in grote trekken gelijk, i.c. een hogere prevalentie voor de kansarmere groepen (de werkelozen, chronisch zieken en de laagst opgeleiden). De duidelijke invloed van de urbanisatiegraad op de prevalentie van psychiatrische stoornissen nam over de periode van 14 jaar (1983-1997) echter af. Leeftijds- of cohort effecten werden niet gevonden, maar een aantal periode effecten zou de stijging in de prevalentie tussen 1983 en 1997 kunnen verklaren. Deze betroffen zowel de populatie in het algemeen (veranderingen in de sociale rolpatronen en in de gezondheidszorgpolitiek) als die van de rurale gebieden in het bijzonder (selectieve migratie vanuit de stad en mogelijk de dreigende overstroming van de huizen en landerijen).
In hoofdstuk 6 (de 'mapping' studie) komt het epidemiologische onderzoek in het kader van het RN-2 project aan de orde. Doel van dit onderzoek was 1) beschrijving van de prevalentie, de socio-demografische spreiding en de comorbiditeit van psychiatrische stoornissen in 1997, gemeten met behulp van de SCAN-2.1 in de open populatie, 2) bepaling van de invloed op de prevalentie, sociodemografische spreiding en comorbiditeit van drie nog niet eerder in de open populatie gemeten diagnostische categorieën (slaap-,

dissociatieve en somatoforme stoornissen) en 3) vergelijking van RN-2 gegevens met de NEMESIS gegevens. De 1-maands prevalentie van psychiatrische stoornissen in de Nijmeegse gezondheidsregio was 17.6% (± 2.6%). Als de drie niet eerder gemeten diagnostische categorieën buiten beschouwing werden gelaten, bleek de prevalentie een derde lager (11.9% ±2.5%). De sociodemografische spreiding liet een significant hogere prevalentie van de psychiatrische stoornissen zien ten opzichte van andere subgroepen voor vrouwen, stedelingen, de leeftijdsgroep 45-54 jaar oud, de lager opgeleide en gescheiden, werkeloze en chronisch zieke personen. De drie niet eerder gemeten diagnostische categorieën hadden geen invloed op deze spreiding. Van alle personen met een psychiatrische stoornis vertoonde 26.1% comorbiditeit. De comorbiditeit van de drie niet eerder gemeten diagnostische categorieën zelf was 36.7%. Als de leeftijdscategorieën van de steekproef en de diagnostische categorieën werden terug gebracht tot de omvang gebruikt in NEMESIS, kwam de prevalentie voor psychiatrische stoornissen in RN-2 met 11.7% (±2.2%) aanzienlijk lager uit dan die gemeten in de NEMESIS studie (16.5%, ±0.4%).
Het gebruik van de SCAN-2.1 met alle beschikbare categorieën van de DSM-IV resulteerde in een prevalentie waaraan de drie niet eerder gemeten diagnostische categorieën een aanzienlijke bijdrage leverden. Het patroon van de sociodemografische spreiding van de psychiatrische stoornissen kwam overeen met de bekende beschrijving van Dohrenwend & Dohrenwend. Ondanks het gebruik van alle beschikbare categorieën uit de DSM-IV en het versterkende effect van de drie niet eerder gemeten diagnostische categorieën, blijkt de gevonden comorbiditeit in RN-2 relatief beperkt. Volledig gestructureerde interviews zoals gebruikt in NEMESIS leveren hogere prevalenties en een hogere comorbiditeit op dan semigestructureerde, klinische, interviews zoals gebruikt in RN-2 .

Deel III

In hoofdstuk 7 staat het onderzoek beschreven naar de psychometrische eigenschappen van de SCAN-2.1 (voor het eerst in volle omvang), dat werd uitgevoerd voorafgaande aan het RN-2 onderzoek. In één onderzoeksopzet werden gegevens vergeleken van gepaarde, onafhankelijke live interviews met dezelfde respondent (test-hertest opzet). In een andere opzet werden tien op video geregistreerde interviews van experts opnieuw gescoord door alle 27 getrainde interviewers (gestandaardiseerde opzet) en de uitkomsten vergeleken met die van de andere interviewers als ook met een referentie score. Voor de test-hertest opzet bleek de kappa coëfficiënt voor de aanwezigheid van ten minste één diagnose aanzienlijk (0.62), voor de te onderscheiden diagnostische categorieën en diagnostische groepen goed tot matig (0.64 tot 0.24). In de

gestandaardiseerde opzet, waarbij tien op video geregistreerde interviews van experts werden beoordeeld, bleken de sensitiviteit en specificiteit aanzienlijk tot bijna perfect. De overeenkomst per interviewer met de referentie diagnoses varieerde van 87% (op niveau van diagnostische groepen) tot 94% (op niveau van het hebben van ten minste één diagnose). Overeenkomst op het niveau van syndromen (zonder duur en hiërarchische criteria van de DSM-IV) was uitmuntend. Hoewel de SCAN meestal wordt gebruikt door clinici, laat dit onderzoek zien dat ook minder ervaren (maar goed getrainde) interviewers het instrument betrouwbaar kunnen hanteren.

Hoofdstuk 8 gaat over de operationalisatie van de SCAN en de effecten van het klinisch oordeel. Omdat het klinisch oordeel is ingebouwd in de scoringsschalen van de SCAN, konden wij de subklinische scores behandelen alsof het klinisch relevante scores betrof. Per saldo telde de subjectief beleefde morbiditeit dan mee in de diagnostische resultaten. Wanneer het klinisch oordeel buiten beschouwing werd gelaten, steeg het aantal positieve gevallen met meer dan één vijfde, van 197 tot 245 personen. De grootste verschillen betroffen somatoforme, psychotische en stemmingsstoornissen. De prevalentie zonder klinisch oordeel kwam uit op 21.4%, 3.8% hoger dan bij gebruik van het formele SCAN-algoritme. Een verhoudingsgewijs gelijke stijging van de prevalentie werd gevonden voor alle sociodemografische groepen. Met klinisch oordeel was een kwart van alle positieve gevallen comorbide, zonder klinisch oordeel steeg dit tot één derde. Het klinisch oordeel had, kortom, op de gevonden prevalentie een aanzienlijk effect, dat groter was bij diagnoses gebaseerd op subjectieve belevingen dan bij diagnoses, waarbij waarneembaar gedrag een overwegende rol speelde. Personen met een diagnose van een lagere orde in de hiërarchie kregen een “belangrijkere” diagnose toegewezen wanneer het klinisch oordeel werd weggelaten. Sommige specifieke diagnoses (zoals somatoforme stoornis niet anderszins omschreven) bleken buitengewoon gevoelig voor het klinisch oordeel.

In hoofdstuk 9 wordt een vergelijking gemaakt tussen het Composite International Diagnostic Interview (CIDI) en de SCAN in de algemene populatie, teneinde beter inzicht te krijgen in de werkwijze van beide instrumenten, die kennelijk uiteenlopende informatie genereren. 139 Personen uit de algemene populatie werden geïnterviewd met zowel de CIDI als de SCAN (in een volgorde compenserende opzet). De volledig gestructureerde CIDI werd afgenomen door goed getrainde leken interviewers, de semi-gestructureerde SCAN door goed getrainde klinisch ervaren interviewers. De overeenkomst tussen de CIDI en de SCAN wat betreft het herkennen van DSM-IV gevallen en specifieke diagnoses was slechts matig. Waar op het niveau van individuele klachten en verschijnselen bij de grote stemmingsstoornissen een slechte tot redelijke overeenkomst werd gevonden, liet een canonische correlatie analyse echter een aanzienlijke tot vrijwel perfecte concordantie zien.

Hoewel de geïdentificeerde latente variabelen van de twee instrumenten een grote samenhang vertoonden, bleken de instrumenten op zich sterk te verschillen wat betreft herkenning van gevallen in het algemeen, specifieke diagnoses en klachten en verschijnselen. Het doel van beide instrumenten om tot dezelfde diagnoses te komen, wordt belemmerd door uiteenlopende operationalisaties van de diagnostische criteria en door verschillende administratieve procedures.

Deel IV

In hoofdstuk 10 worden de empirische bevindingen uit deel II en III beschouwd vanuit het theoretisch perspectief van de psychiatrische epidemiologie en het diagnostische proces (deel I). Een en ander resulteert in aanbevelingen gericht op de verbetering van het diagnostische proces in de psychiatrische epidemiologie. Allerlei bronnen van diagnostische variantie blijken de vergelijking van ruwe prevalentie maten binnen de descriptieve epidemiologie te compliceren. Van belang zijn met name de classificatie systemen en interview strategieën, omdat deze naar de achtergrond verdwijnen c.q. impliciet worden in de fase dat prevalenties worden berekend en gerapporteerd. Met betrekking tot de diagnostiek kunnen we concluderen dat het semigestructureerde, op klinisch oordeel gebaseerde SCAN-2.1 interview bruikbaar is voor gevalsidentificatie in de open populatie en dat de toepassing van dat klinische oordeel belangrijke effecten heeft op die gevalsidentificatie. Het klinisch oordeel lijkt dan ook een onmisbaar instrument om valide informatie te verzamelen omtrent de stoornissen in psychiatrisch epidemiologisch bevolkingsonderzoek.
Tot slot bevelen wij aan het classificatie systeem en de interview strategie als expliciet referentiekader te betrekken bij de waardering van psychiatrisch epidemiologische onderzoeksresultaten en daaraan conclusies te verbinden vanuit klinisch perspectief.

Bijlage

De bijlage bevat twee delen.
Het eerste deel betreft het gepubliceerde cross-sectionele RN-onderzoek gebaseerd op de GHQ gegevens, een onderdeel van het RN-project dat als referentie is gebruikt in hoofdstuk 5.
Het tweede deel bestaat uit de conversietabel van de SCAN/PSE-10 scores en de PSE-9 scores.

Dankwoord

Het behoeft geen betoog dat het schrijven van een proefschrift slechts met goed gevolg kan worden afgerond als er vele sterke schouders zijn die dat proces ondersteunen. In de ruim 11 jaren die zijn verstreken sinds de allereerste overwegingen om in het Nijmeegse Regioproject te stappen, zijn dit er zoveel dat ik me moet beperken tot die personen, die mij veelvuldig en nadrukkelijk in deze lange periode hebben bijgestaan.

Ernst Roscam Abbing, Bert Tax en Christiane König (zij mag deze dag helaas niet meer meemaken) wil ik als medewerkers van het eerste uur hartelijk danken. Ook hartelijk dank aan de andere medewerkers van Sociale Geneeskunde, maar speciaal aan mijn maatjes van 'de zaal': Josien van den Berg, de eerste jaren altijd aanwezig als ik er maar twee dagen in de week kon zijn; Saskia Zandstra, met een niet aflatende ijver om de gegevens van huisartsen te verzamelen en later Eric van Rijswijk en Nathalie Donders. Jan Mulder wil ik danken voor zijn altijd weer indrukwekkende 'simpele' kijk op wat er statistisch moest en kon gebeuren. Fokko Nienhuis en Rob Giel wil ik hartelijk danken voor de introductie in de SCAN die ze hebben verzorgd Alle SCAN-interviewers zonder wie het betrouwbaarheidsonderzoek nooit tot een goed einde was gekomen en die daarnaast samen goed zijn geweest voor 983 veldwerk interviews: hartelijk dank.

De GGz Midden Brabant (en speciaal Willem van Hezewijk en Gerard Coonen) wil ik hartelijk danken voor de mogelijkheid die zij mij hebben geboden om het werk in Nijmegen nog een tijd door te zetten, nadat de initiële subsidie was opgesoupeerd. Ook Marlies Geurts wil ik hartelijk danken voor haar steun en ordening op de achtergrond, waardoor de andere werkzaamheden gewoon konden doorgaan en mijn aandacht bij het afronden van het proefschrift kon blijven.

Paul Hodiamont, mijn promotor, en de aanstoker van dit alles. In Utrecht op een terras aan de Gracht bij een Cola kwam het eerste idee, in Atlanta met uitkijk op de Coca Cola toren ontstonden de eerste teksten. Maar zeker ook tijdens het verder volgen van het schrijfproces heb ik (naast dat ik het er wel eens moeilijk mee had) zijn steun en kennis enorm gewaardeerd.

Joop Furer was altijd een grote steun in moeilijke momenten (waren het maar momenten?) en van onschatbare waarde bij de kritische analyse van mijn teksten. Zijn overtuiging dat het allemaal nog goed zou komen heeft mede gemaakt dat dit proefschrift tot een goed einde is gekomen.

De overige leden van de promotiecommissie, Aartjan Beekman, Harm Gijsman, Guus van Heck, Koos van der Velden, Jolanda de Vries en Frans Zitman, dank ik voor de tijd en de aandacht die zij hebben besteed aan dit proefschrift.

Dank aan Renier van Dinter van Janssen-Cilag en een speciaal woord van dank aan Jeff Lubin, voor zijn vele uren en de prettige samenwerking waarin hij mijn Engels aan Britse maatstaven onderwierp.
Ik dank AstraZeneca, Bristol-Myers Squibb, GlaxoSmithKline, Janssen-Cilag, Lilly, Lundbeck en Wyeth voor hun bijdrage in de kosten van de uitgave van mijn proefschrift.

Ik ben ook veel dank verschuldigd aan mijn meest dierbaren. Mijn vader en Marijke dank ik hartelijk voor de vele jaren waarin zij hun belangstelling hebben getoond. Mijn vader maakt dit niet meer bewust mee, maar in gedachten zie ik hem zitten, en glunderen.
Ralph en Karen, het is nu eindelijk gedaan. Papa komt weer eens achter zijn laptop vandaan.
Tot slot wil ik Claire heel hartelijk danken voor die vele jaren van ondersteuning, waarin zij van ons huis een thuis bleef maken en voor alle kritische noten, al waren die soms moeilijk te kraken.

Curriculum Vitae

Cees A.Th. Rijnders was born in Tilburg on February 21, 1956. He attended secondary school in Tilburg (Paulus Lyceum) and Breda (Newman College), which he finished in 1974. He studied Medicine at the Erasmus University Medical School of Rotterdam as from 1974 and received his medical degree in 1981. Subsequently, he worked as a neurological resident at the Bethel Hospital in Delft and started his specialisation as a psychiatrist in 1983 at the Jelgersma Clinic, Endegeest and the Leiden University Medical Center (director of residency training Professor dr. H.G.M. Rooijmans). In 1987 he was registered as a psychiatrist after a last year of residency training at the Regional Institute for Ambulatory Mental Health Care (RIAGG) Westhage in the Hague.
From 1987 until 1997 he worked as the head of the outreaching emergency treatment centre of the three RIAGGs of the Hague and as psychiatrist in the Crisis Intervention Centre. In 1997 he was appointed as a director of residency training at the Mental Health Care Institution GGZ-Midden Brabant in Tilburg and he started to work on the Nijmegen Health Area project.

Notes

[i] The disease, the illness and the sickness model are explained in chapter 3.

[ii] When reanalysed in accordance with the 1997 procedure, the 1983 psychiatric prevalence rate turned out to be 7.8% (±2.3%).

[iii] Psychosomatic problems (item 3), delusional anxiety (item 13), pre-menstrual increase of symptoms (item 39), delusions concerning special guidance (item 75), denial or concealment of delusions and hallucinations (item 94), and insight into neurotic disorders (item 105)

[iv] (±95% confidence interval)

[v] the Short Form Health Survey (SF-36), 4-neurotic symptoms (4-NS) (Henderson et al. 1981; Vermeulen & Bosma 1992), Social Support List 12-item version (SSL 12-I), the short version of the Groningen Questionnaire about Social Behavior (GQSB), Loneliness Scale (de Jong-Gierveld & van Tilburg 1993), Brugha Life Events Questionnaire (Brugha et al. 1985), short version of the Coping Inventory for Stressful Situations (CISS-21) and Self-Reported Chronic Diseases (Mootz & van den Berg 1989; van den Berg & van den Bos 1989), psychiatric history, modified from the Longitudinal Aging Study Amsterdam (LASA), and some questions on political and religious convictions.

[vi] This exception will be specified in 7.3.3.2

[vii] DSM-III-R criterion A4 (frequent intoxications or withdrawal symptoms when the individual is expected to fulfil major role obligations at work school or home) has been removed and the withdrawal criteria A8 and A9 of the DSM-III-R have been combined into a single criterion in the DSM-IV.

[viii] DIS: Diagnostic Interview Schedule; CIDI: Composite International Diagnostic Interview

[ix] PSE: Present State Examination; SCAN: Schedules for Clinical Assessment in Neuropsychiatry

[x] 'not known' is scored 8 and 'uncertain whether present or not' is scored 9

[xi] For example, a subject is very concerned about something that the interviewer assesses as subclinical or irrelevant, or the subject presents real symptoms, but they are assessed as situationally adequate for the particular subject.

[xii] Landis and Koch (1977) proposed describing the degree of concordance as follows: < 0.21, “poor”; 0.21-0.40, “fair”; 0.41-0.60, “moderate”; 0.61-0.80, “substantial”; and 0.81-1.00, “almost perfect”.

[xiii] For practical purposes, we formulated the following definitions:

- **case definition**: the theoretical concept underlying ‘what is a case’, based on the presence of a specified set of signs and symptoms. Case definitions are ordered in a classification system, on the one hand by describing the boundaries of each case and thus the distinctness of each case with reference to all others, and on the other hand by sorting cases into categories through matching characteristics;
- **case identification**: the operationalisation of the theoretical diagnostic concept or classification system on the basis of a detailed description;
- **case finding**: the application of case identification in a population study.

[xiv] Standardised interview: the interview is conducted on the basis of a fixed question schedule which is determined in advance;
Semi-structured: the standardised question schedule is composed of compulsory questions, followed by optional questions allowing the interviewer to probe for additional information until satisfied. A bottom-up approach is used in the interview in a trawl-like search for positive signs and symptoms. As a consequence, when the classification system changes, the interview itself can be retained; only the algorithm for establishing the diagnoses has to be rewritten.
Fully structured: the standardised question-schedule is rigid and compulsory, without the opportunity for individual probing. The interview is based on a clear psychiatric classification system and is consequently restricted to questions that give a direct answer as to the question of whether certain diagnosis-related criteria are fulfilled (top-down procedure). As a result, when the classification system changes, both the interview itself and the algorithm for establishing the diagnoses have to be rewritten.

[xv] ICD: International Classification of Diseases

[xvi] DSM: Diagnostic and statistical manual of mental disorders

[xvii] The Composite International Diagnostic Interview (CIDI), written at the request of the World Health Organization/US Alcohol, Drug Abuse, and Mental Health Administration Task Force on Psychiatric Assessment Instruments, combines questions from the Diagnostic Interview Schedule with questions designed to elicit Present State Examination items (Robins et al. 1988).

[xviii] Version of July 1997

[xix] 18 to 64 years of age, as used in the NHA-1 study and in the NEMESIS

[xx] As used in the NEMESIS study (and highly similar to the NHA-1 study, but with DSM-IV instead of ICD-9 for classification)

[xxi] Indicative clinical severity was based on the lifetime history of suicidal ideation and suicide attempts and on whether in the past month the respondent had been unable to work or had to cut back on work or usual activities for two or more days.

[xxii] Mostly anxiety and depression (Fryers et al. 2005)

[xxiii] 1-month comorbidity data from the NCS and the NCS-R study are not available (Kessler et al. 1994; Kessler et al. 2005a; Kessler et al. 2005b).

[xxiv] There is a broad variety of definitions for comorbidity, which results in an array of operational applications and an extended range of comorbidity rates (van den Akker et al. 1996). When the issue is restricted to psychiatric comorbidity, a major question is the reference period during which the disorders should co-occur. Feinstein (1970) defines comorbidity as “a distinct additional entity that has existed or may occur during the clinical course of a patient who has the index disease under study”. By definition, comorbidity in studies with point (or 1-month) prevalence rates is operationalised as two or more distinct entities existing at the same time (concurrent comorbidity). In studies aiming at longer reference periods, like a 1-year or a lifetime prevalence rate, there may be two or more disorders during the period of reference, not only at the same time (concurrent comorbidity) but also successively (consecutive morbidity). Because comorbidity in studies aiming at a prolonged reference period is used in both senses, the comorbidity rate tends to be higher. The effect of mixing

concurrent and consecutive comorbidity is illustrated by the results of the NCS and the German National Health Interview and Examination Survey (and with respect to the 1-year and lifetime reference periods, also the NEMESIS). With the prolongation of the reference period, comorbidity rates in subjects with at least one diagnosis increased from around 30% for the 1-month period to over 50% when lifetime was used as a reference period (Bijl et al. 1998; Jacobi et al. 2004; Kessler et al. 1994; Ravelli et al. 1998).

[xxv] The transition from DSM-II to DSM-III and its successors led to a growing number of specified diagnoses, accompanied by a recession of the hierarchical approach and has inevitably brought about an increase in the prevalence rate of comorbid psychiatric disorders (Kessler et al. 1997; Pincus et al. 2004). For example, the number of DSM-categories increased from 127 in DSM-I (1952) through 226 in DSM-III (1980) to 326 in DSM-IV (1994).

From this perspective, an important possible explanation of the considerable comorbidity rates was advanced by Maj (2005), who stated that to his knowledge the rule according to which the same symptom can not appear in more than one disorder has never been made explicit in DSM-based publications, and this rule creates an artificial delineation of syndromes. Pincus (2004) argues that the next evolution of the DSM needs to balance the current 'rule-based system with diagnostic strategies that depend on clinical judgement'.

Another artefact has arisen through the use of the term comorbidity to indicate subjects presenting complex clinical presentations that do not meet the typological classification criteria of a single diagnosis, resulting in a comorbid classification (Alonso & Lepine 2007; Pincus et al. 2004). This artificial splitting up is one of the reasons for the morbid growth of the number of diagnoses in the DSM-IV, which has led to the resurrection of the hierarchy rules. Although the influence on the comorbidity rate of switching off the hierarchical rules in DSM-IV was only minimal in NEMESIS (Ravelli et al. 1998), the comorbidity rate for subjects with at least one diagnosis in the NHA-2 increased from 26% to almost 43% (Rijnders & Furer 2003). The explanation for this finding is the fact that in the NHA-2 study, three diagnostic categories were measured with an increased contribution to comorbidity that were not part of the diagnostic range in the NEMESIS. In our study, the hierarchy rules were used explicitly and studied for their effect (see sections 6.3.3.2 and 6.4.5). This procedure resulted in more moderate comorbidity rates. The highest

comorbidity rate was found when the hierarchical rules were omitted (Table 1).

Table 1 Prevalence rates and associated comorbidity rates with respect to diagnostic range, hierarchy and clinical judgement

	Prevalence rate	Comorbidity rate
Original version, with hierarchy and clinical judgement	17.6	26.1
Without hierarchy	17.6	42.5
Omission of clinical judgement	21.4	31.6

A more moderate increase was found when clinical judgement was omitted. As reported in chapter 8 on the clinical judgement of the SCAN, this increase was largely due to an increase in the prevalence rate of the three newly measured diagnostic categories. A reduction of the diagnostic range whereby the diagnostic categories lower in the hierarchy were omitted had a considerable effect on both the prevalence rate and the comorbidity rate (11.9% and 22.3%, respectively).

[xxvi] In general, fully structured interviews generate higher diagnostic prevalence rates than do semi-structured interviews. Theoretically, when independence of the diagnostic categories is adopted as a premise, higher prevalence rates lead to a higher minimum comorbidity rate. When, as argued earlier, the diagnostic categories are not entirely separate, independent entities, the actual comorbidity rate rises above this theoretical minimum rate. Moreover, fully structured interviews rely on perceived morbidity of signs and symptoms without a clinical attribution to a certain diagnosis. This subjectivity may lead to the attribution of one and the same sign or symptom to more than one diagnosis and consequently it artificially increases the comorbidity rate.

[xxvii] Healthy subjects, patients seen by GPs in their surgery, outpatients and inpatients identified by mental health care

[xxviii] Although Brugha et al. (1999b) reported data suggesting ‘progress in overcoming the persistent problem of lay rater bias identified in previous studies’ (Rodgers & Mann 1986), this pronouncement is exaggerated. These investigators did not use average, but highly experienced lay interviewers who followed a considerably extended training period in clinical evaluation. Furthermore, the positive cases were recruited from a

hospital site and consequently were too easy to identify because of the presence of obvious signs and symptoms.

[xxix] For the clinical practice situation, the Mini-SCAN, a completely computerised condensed version of SCAN, has been developed and should be the clinical interview instrument of choice for use in daily practice (Nienhuis & Giel 2006).

[xxx] When the interviewer is convinced that the remaining part of the section will not yield any major symptoms relevant to the assessment of a possible diagnosis, the rest of the section is omitted.

[xxxi] For example, the items necessary for dysthymia are obtained by reviewing items on depression that were assessed earlier in the section.

[xxxii] At the conversion from SCAN to PSE-9, on this item the lower of the two possible scores has been chosen. Starting from SCAN, basically two scores (2 and 3) in PSE-9 are possible. Score “2” seems the most in accordance with the qualitative intention.

[xxxiii] Does not discriminate between PSE-9 question 47 and 48. For CATEGO this is not of importance according to WHO-committee.

[xxxiv] Does not discriminate between PSE-9 question 47 and 48. For CATEGO this is not of importance according to WHO-committee.

[xxxv] The PSE-9 scores 1 and 2 are not separately convertible from SCAN. The highest PSE-9 score has been chosen for not losing positive signs in the conversion process.

[xxxvi] The PSE-9 scores 1 and 2 are not separately convertible from SCAN. The highest PSE-9 score has been chosen for not losing positive signs in the conversion process.

[xxxvii] Because de final version of Section 21 of SCAN-2.1 was not yet available at the time of the training and reliability study, the decision was made for the fieldwork to use the trained version of SCAN-2.1 of spring 1997, The item numbers of section 21 have been adjusted.