MENTAL HEALTH IN DUTCH CHILDREN

an epidemiological study

Cover: Thea van Doorn

ISBN 90-9001001-7

From the department of Child Psychiatry, Academic Hospital Rotterdam - Sophia Children's Hospital Rotterdam, The Netherlands.

MENTAL HEALTH IN DUTCH CHILDREN

an epidemiological study

(Psychische gezondheid bij Nederlandse kinderen
een epidemiologisch onderzoek)

Proefschrift

ter verkrijging van de graad van doctor in de
geneeskunde
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof. Dr. M. W. van Hof
en volgens besluit van het College van Dekanen.
De openbare verdediging zal plaatsvinden op
woensdag 18 september 1985 te 15.45 uur

door

FRANK CORNELIS VERHULST

geboren te Rotterdam

krips repro meppel

Begeleidingscommissie:

Promotoren: Prof. Dr. J. A. R. Sanders-Woudstra
Prof. Dr. H. A. Valkenburg

Overige leden: Prof. T. M. Achenbach, Ph.D.
Prof. Dr. W. J. Schudel

This research was made possible by financial support from the Sophia Fund for Scientific Research.

To Mia
Francien, Susanne and
Paul

TABLE OF CONTENTS

ABBREVIATIONS

CAS	Child Assessment Schedule
CBCL	Child Behavior Checklist
DSM-III	Diagnostic and Statistical Manual of Mental Disorders
ICC	Intraclass correlation coefficient
SES	Socio-economic status
TRF	Teacher Report Form
WISC-R	Wechsler Intelligence Scale for Children - Revised edition
ANCOVA	Analysis of Covariance
ANOVA	Analaysis of Variance

CHAPTER I
INTRODUCTION

I.1 Background and motivation

Epidemiology is the study of the distribution of a disease or a physiological condition in human populations and of the factors that influence this distribution (Lilienfeld, 1976).

Since World War I epidemiology has gradually shifted its emphasis from the study of the distribution and causation of infectious diseases to the study of nutritional deficiencies such as pellagra and still later to the inquiry into various important chronic diseases. This broadening scope of epidemiology is reflected in its definition by Cassell (1965) who states that epidemiology is "one of the sciences concerned with the study of the processes which determine or influence the physical, mental and social health of people. It is with their health in relation to their behaviour in social groups that epidemiology is primarily concerned". This definition, reflecting a paradigm influenced by sociologic thought, is certainly not shared by all epidemiologists. However, it reflects a certain change from mono-causal disease models towards a multicausal model in which pathogenic agent, the personal characteristics of the diseased subject and the environment are mutually influencing each other.

As epidemiology evolved and consequently its concepts and methods were adapted for chronic and presumably non-infectious diseases, such as cancer, high blood pressure and coronary artery disease, it became applicable to psychiatric disorders. Most "modern" diseases and psychiatric disorders have in common that they are chronic and multifactorially determined. Over the last 25 years epidemiology has been introduced into child psychiatry. Although in the preceding years quasi-epidemiological studies were reported on the prevalence of behavior problems in non-representative samples of children, it lasted until 1958 when Lapouse and Monk published the first report of the frequency of behavior problems in a representative sample of randomly selected mothers of 6-12-year-olds. This study was the first true child psychiatric epidemiologic enterprise undertaken so far.The epidemiological approach added a new dimension to the understanding of psychiatric disorders in childhood, which until then had been derived largely from individual case-studies and not from the study of problems as they occur in groups of children in a total population.

Since Lapouse and Monk's (1958) study, a number of community surveys have been carried out, of which the Isle of Wight study (Rutter e.a. 1970), has been the most influential because of its sound methodology and its combination of statistical and clinical approaches. In this survey the total population of over 3000 10-and 11-year-old children was studied

initially. Until Richman e.a. (1975) published their survey of behavior problems and developmental delays in a total population of 950 3-year-old children, no data had been available on behavior problems in representative samples of preschool children. The authors reexamined more than 90% of the initially studied children five years later and arrived at the important conclusion that 61% of the problematic 3-year-olds were still showing significant difficulties five years later (Richman e.a., 1982).

Most studies of the epidemiology of behavior problems have focussed on a narrow range of behavior and/or age. Achenbach e.a. (1981), however, provided prevalence data on behavioral problems and competencies of 1300 children of 4-16 years old who had been referred to mental health services. The authors compared the referred sample with data on 1300 nonreferred children well matched for age, gender, socio-economic-status (SES), and race. Furthermore, these data were used as a basis for developing instruments (Achenbach, 1978) useful to epidemiological and other research.

Although progress has been made over the last twenty-five years, child psychiatric epidemiology struggles with many limitations mainly caused by the lack of accepted definitions of the disorders and by the nature of the disorders themselves. Except for a few rather rare conditions such as infantile autism, Tourrette's disorder and a few other syndromes, child psychiatric disorders do not consist of clearcut disease-like entities. Most behaviors that constitute the majority of child psychiatric disorders also occur in normal children. Therefore, the greatest difficulty is the determination at which point we consider behavior, or a combination of behaviors, as pathological. As most child psychopathologic conditions lack accepted definitions as to the degree and type of disorder, one may wonder how child psychiatric epidemiology is ever possible without consensus on definitions. However, epidemiologic research may in fact be an important step in arriving at valid definitions (Achenbach, 1980 a). For instance, one of the findings of the Lapouse and Monk (1958) study was that for a substiantial number (43%) of normal children, seven or more fears were reported. This finding stressed the need to reconsider the criteria by which anxiety phenomena should be called normal or abnormal.

Another important field in which the reciprocity of the relationship with epidemiology plays a relevant role is the construction, testing and interpreting of instruments for measuring degree and type of psychopathology. Epidemiology of child psychiatric disorders cannot be carried out without instruments for measuring behavior or guidelines in order to make judgements about the behavior studied, but on the other hand epidemiological methods to study the normal distribution of types of behavior is crucial to construct and interpret tests. This does not apply only to the measurement of the child's behavior. The same principles apply to intelligence tests, for instance, but also to the interpretations of measurements of physical phenomena such as blood-pressure.

A standardized measure of child behavior problems is the Child Behavior Checklist (CBCL). The CBCL, developed by Achenbach (1966,1978), is a checklist designed to collect data on a wide variety of behaviors that are of clinical concern. The CBCL is filled out by the parents of the child.

Furthermore, data obtained with the CBCL can be scored on the Child Behavior Profile, a standardized profile for portraying behavioral disorders and competencies. Using epidemiological methods, the CBCL was normed for the American population of children 4 through 16 years old (Achenbach e.a., 1981).

At the time the present study was started (1982) correlations were not available between CBCL scores obtained from parents and independently obtained information on the same child through a teacher questionnaire, nor between the CBCL and intensive clinical assessment of the same child.

Although the number of publications on child psychiatric epidemiology has increased over the last ten years, there are very few published accounts of investigations which attempt to replicate and refine existing techniques in populations different from those for which the techniques were originally developed. Earls (1980) stresses that trans-cultural attempts at replicating results with the same method represent an essential means of assessing the extent to which features of children's behavior disorders are universal.

At the time the present study was started, except for studies concerned with isolated symptoms such as sleep problems (Cohen-Matthijsen, 1980) and enuresis (De Jonge, 1969), there were no data available in the Netherlands on the prevalence and distribution of child mental health problems of a wide age and behavior range in representative population samples. The availability of outpatient and inpatient child mental health facilities in this country is in general arbitrarily determined and is not based on population data on the prevalence and distribution of behavioral, emotional and developmental problems. Furthermore we were confronted with the fact that virtually no validated instruments for determining degree and type of child psychiatric disorder in the Dutch population were available. This serious lack of basic research tools was not only felt as an obstacle for epidemiologic research. Therapy-evaluation and etiological studies were also hampered by this situation.

I.2 Aims of the study

In the situation described above as a starting point, there was ample justification for designing a study with the following general aims:

1. to provide prevalence data on specific emotional and behavioral problems in a sample of the Dutch population of children aged 4-16.
2. to provide prevalence data of overall psychiatric disorder in children in a sample of the Dutch population of children aged 8 and 11.
3. to identify differences related to demographic and other clinically relevant variables.
4. to compare Dutch data with data from other surveys reported in the literature.
5. to relate data obtained by an already existing instrument for describing relevant child behavior as reported by parents, the Child Behavior Checklist (CBCL) (Achenbach, 1978), to other assessment procedures.

I.3 The structure of the report

In chapter II, basic concepts and methods in child psychiatric epidemiology are dealt with, and the results of relevant child psychiatric prevalence studies are reviewed.

The actual study consisted of two stages. In the first stage, data were obtained on specific items. Parents of 2076 children aged 4-16 from the general population completed the Child Behavior Checklist (CBCL). For 1179 children aged 4-12 of the main sample, teachers completed a questionnaire (the Teacher Report Form, TRF) analogous to the CBCL. In addition, for 1387 children newly referred to child mental health settings, CBCL's were obtained. The methods used in this first stage are described in chapter III. In chapter IV, the results of the first stage will be presented and in chapter V the findings are discussed.

In chapter VI the methods employed in the second stage of the study are described. In this second stage, 8-and 11-year old children were selected for intensive interviewing if their scores on the CBCL and/or on the TRF indicated that they showed behavior problems. Chapter VII deals with the results of the second stage, and in chapter VIII these results are discussed. In chapter IX the total study is evaluated.

CHAPTER II
REVIEW OF THE LITERATURE

In this chapter the most relevant concepts and methods in the epidemiology of child psychiatric conditions are dealt with and major results reported in the literature are reviewed.

II.1 Aims of child psychiatric epidemiology

The epidemiology of child psychiatric disorders consists of two major branches, both concerned with studying child mental health problems in the community as a whole or in representative samples. The first branch is especially concerned with understanding the etiology of psychiatric disorders, and is called etiologic epidemiology. The second branch, called administrative epidemiology, is needed for answering questions that are of importance for child mental health planning. The epidemiology of child mental health problems serves a number of important purposes. These include:

1. The determination of the prevalence and duration of child mental health problems in the community.
2. The determination of the association of demographic variables such as gender, age, socio-economic status and ethnicity with child mental health problems.
3. The quantification of the need for services for child mental

health problems.
4. Evaluation of the functioning of mental health services for children.
5. The prevention of child mental health poblems.
6. The investigation of historical trends in child mental health problems and the isolation of factors influencing their development in a favorable or unfavorable direction.
7. The search for etiologies of child mental health problems.
8. The generating of hypotheses for further research.
9. The providing of base-line data for other investigations.

II.2 Definition of child psychiatric disorder

The basic unit of study in epidemiology is the identified case in the population. However, as we have already discussed in paragraph I.1, there is no consensus on what can be called a child psychiatric disorder. As most child psychiatric disorders do not consist of disease entities and most behavioral difficulties also occur in otherwise normal functioning children, the major difficulty in defining child psychiatric disorder lies in the decision on how and where to place the area between normality and pathology.

Three important factors inherent to the disorders themselves hamper consensus on definition: the role of development, the situational specificity of the child's behavior, and the issue of illness-recognition.

II.2.1 The developmental perspective

Most specific behavioral difficulties, for instance temper tantrums, or patterns of behavioral difficulties like school refusal, can be judged as normal at one age whereas they will be labeled as abnormal at another age. For example a 3-year-old child desperately protesting temporary separation from his mother in the waiting room shows a normal phenomenon, whereas the same behavior in a 13-year-old child would be judged as an indication of psychic disorder. In judging whether a child's behavior is deviant or not, one has to take account of the developmental level of the child. Therefore, knowledge is needed on what kind of behavior is normal for different ages. Prevalence estimates of a variety of clinically significant behaviors or behavior-complexes at different ages in the general population offer such knowledge.

The impact of behavioral difficulties on the child's development is an important determinant of pathology. School refusal, for instance, can be a sign of brief emotional upset. However, when the child stays at home instead of going to school for a prolonged period this can interfere with further development.

II.2.2 Situational specificity of behavior

The child's behavior often varies greatly with the situation where the

behavior is observed. Hyperactivity, for instance, may be confined to the school situation only (Klein e.a., 1975), whereas disobedience may be a problem at home but not at school. The same is true for behavior observed in the office or play-room. The interaction of the child with his immediate surroundings plays a central role in the emergence of problems.

II.2.3 Illness recognition

Whereas adults are able to decide whether their symptoms are worthy of professional attention, children seldom seek help for their problems on their own. They are taken by their parents to mental health services, mostly because the parents detect some problem in their child or because the schoolteacher or school official advises help. In the adult it is his own illness recognition which is crucial to his seeking help (Goldberg, 1980). In the child, however, it is the perception of others which decides whether help is sought or not. Sheperd e.a. (1966) found evidence that referral to a child guidance clinic is related chiefly to parental reactions and not to the child's behavior disturbance. Mothers of clinic children were more apt to be anxious, depressed and easily upset by stress. They were less able to cope with their children, more apt to discuss their problems and to seek advice as compared to mothers of children from the normal population. However, in this study the clinic group was selected after the most serious cases were excluded.

Achenbach e.a. (1981) stated that actual referral status is probably as good a morbidity criterion as any other, because actual referral reflects persisting problems on the part of the child in one or more important life areas. The authors found in their survey, which was well controlled for age, gender, SES and race, that 116 of the 118 behavior problem items studied were significantly associated with referral status. It must be stressed that, although this study supports the use of referral status as a morbidity criterion, this does not mean that referral status can be used as the only source of case detection, because many children in the general population showing psychiatric problems do not receive professional help (Rutter e.a., 1970; Sheperd e.a., 1966).

II.2.4 Practical approaches

In order to define a psychiatric disorder several approaches have been used in surveys designed to determine the prevalence of child psychiatric disorders in the general population.

II.2.4.1 Individual symptom approach

At the level of individual items or symptoms of behavior known to be of clinical importance, authors (Lapouse e.a., 1958; Rutter e.a., 1970) agree on the limited value of most individual symptoms as a morbidity criterion for overall functioning. The authors arrive at this conclusion because many symptoms observed in children referred for psychiatric help, such as fears,

occur with a high frequency in children from the general population. This does not automatically mean that all individual symptoms are insignificant to child psychiatrists. Indeed, it is an important fact that many children with behavioral and/or emotional difficulties are not seen by psychiatrists. But does the fact that the majority of children with, for example, a skinrash are not seen by pediatricians imply that pediatricians or epidemiologists should not take an interest in this symptom ? Certainly not, and the same is true for behavioral difficulties.Symptoms like bedwetting, constipation, stuttering, sleepwalking and nightmares have only a weak relationship, if any, with overall psychiatric dysfunctioning (Achenbach e.a., 1981). Nevertheless, child psychiatrists are confronted with children who exhibit these symptoms and who are referred for assessment and treatment.

A number of individual symptoms are so rare in the general population that any child of a certain age exhibiting this symptom can be regarded as disordered. Encopresis after the age of 5, eating non-food, public masturbation or daytime wetting after the age of 7, hallucinations, suicide attempts, and truancy in young children are a few examples of clinically significant symptoms which occur rarely in the general population and which are sometimes even pathognomonic for certain psychiatric conditions.

II.2.4.2 Item count approach

The second approach towards defining psychiatric disorder is the summing up of scores obtained across all items on a questionnaire or assessment instrument. This sum of scores is called the total score. Using this approach, Glidewell e.a. (1957) showed that, in a sample of 91 schoolchildren and 35 clinic cases there was a positive relationship between the number, frequency, duration and severity of the symptoms reported by a child's mother and the degree of sickness found in the child by independent assessment. The subject's total scores can be handled statistically either as an interval scale or, by selecting a cutoff point, as a nominal scale. In other words, the total scores on a parent questionnaire can either be rank ordered from lowest to highest, reflecting children from least to most disordered, or the total scores can be used to categorize children into normal and disordered groups by selecting a specific score above which children are called disturbed. As most psychopathological disorders in childhood can be regarded as quantitative rather than qualitative deviations, the sum of scores on a questionnaire designed to cover the child's behavioral problems can be regarded as a dimensional measure of behavioral problems analogous to total scores on an intelligence test representing general ability as a construct. Most child psychiatrists prefer dichotomising measures of behavioral deviations into groups representing disordered and non-disordered children. This approach has the advantage of clarity and surveyability, but, the replacement of dimensional measures by categorical ones has the disadvantage of discarding valuable statistical information.

In three studies, selected for their influence on later work, (Achenbach e.a., 1981; Richman e.a., 1982; Rutter e.a., 1970), the critical or cutoff score defining deviance on the screening instrument was determined by comparing clinic versus non-clinic attenders. The cutoff scores were chosen on the basis of their discriminative power. The cutoff score should ideally be correctly predicting non-cases as normal and cases as disordered. Two measures reflecting the discriminative power of an instrument are the sensitivity and the specificity. Sensitivity is the percentage of cases in a population correctly identified as cases. The specificity is the percentage of non-cases correctly identified as non-cases. By subtracting these percentages from 100 the percentage of false-negatives and false-positives respectively is found. Achenbach e.a. (1981) arrived at a total misclassification rate (number of cases incorrectly identified by the questionnaire as a percentage of the total number of patients screened) of 17.6% for the 90th percentile of total behavior problem scores from the Child Behavior Checklist (CBCL). This misclassification rate dropped to 15.5% when social competence scores below the 10th percentile was added as a criterion. Rutter e.a. (1970) found a total misclassification rate of 19.2% on their "children's behaviour questionnaire" for completion by parents using a cutoff point of 13, while Richman e.a. (1971) in their survey of behavior problems in 3-year-old children found 25% false negatives and 13.1% false positives.

However, as outlined in paragraph II.2.3, it could be argued that the referral of a child to a mental health service does not always imply that the child has a mental disorder. Marital problems, disputes over custody, and problems with school achievement, are examples of other reasons for referral to child mental health services. Furthermore the choice of the population of referred children studied can be a source of bias. Richman e.a. (1971) obtained information on only 20 highly selected clinic attenders, whereas Achenbach e.a. (1981) obtained data on 1300 children attending mental health facilities in a large geographic area.

Another approach to determine the cutoff score above which mental disorder is supposed to be present is to compare the questionnaire's total score with psychiatric judgement. Rutter e.a. (1970) compared psychiatrists' overall impairment ratings based on information obtained through parent interviews, direct psychiatric assessment of the child and on teachers reports, with children's total scores on both parent and teacher questionnaires. On the basis of their results, the authors did not recommend adjusting for the cutoff levels of the questionnaires. Richman e.a. (1982) used the "Behaviour Screening Questionnaire" in order to identify 3-year-old children with behavior problems in the general population. Psychiatrists gave each child a clinical rating of disturbance based on all information about the child's behavior gathered by the interviewer who administered the questionnaire to the mother during a home interview. The authors used these clinical ratings as a check whether the cases identified by the screening questionnaire were really disturbed. Following this method they found a total misclassification rate of only 8.6%.

Another study in which questionnaire scores were compared with overall impairment ratings and direct clinical assessment is the Langner e.a. (1976) study. A 654 item questionnaire was administered to a cross-sectional sample of 1034 mothers from the general Manhattan population and to a welfare sample of 1000 mothers of children between the ages of 6 and 18. All the questionnaires were rated by psychiatrists on a 5-point scale of total psychiatric impairment. Thirty-five items of the total 654 were selected on the basis of their high correlation with the psychiatrists' total impairment ratings. In this study the total behavior problem score was not used for categorizing children into disturbed and non-disturbed groups. However, the total scores were used as an interval scale which had the advantage of retaining more potentially relevant statistical information. In a sample of 271 children, the 35-item total score was found to correlate .69 with total impairment rating, .33 with direct psychiatric assessment of the child and .49 with lifetime treatment referral status. The correlation coefficient between referral status and direct psychiatric assessment was only .29. As in the Richman e.a. (1982) study, Langner e.a. (1976) based their impairment ratings for the greater part on information from the questionnaire itself. Although this procedure was treated as a check on whether the questionnaire's total score really indicated the degree of mental disturbance, it certainly was not an independent check, because it was based on information from the questionnaire itself. Psychiatrists making impairment judgements were likely to be influenced by the number of items, which was also reflected in the total score. From the Langner e.a. (1976) study it can be concluded that referral status as an independent external check had the highest correlation with total behavior problem score.

An argument against the use of simply summing up all the scores obtained across items on an instrument is the fact that the items are entered without being weighed. For instance an item like "nailbiting" contributes to the total score in the same degree as an item like "talks about killing self". One statistical approach for handling this problem is discriminant analysis. Achenbach e.a. (1981) used this method for discriminating between their non-referred and referred groups. The essence of discriminant analysis, a multivariate statistical approach, is the provision of optimal discrimination between previously identified populations (here, referred and non-referred children) by weighing items in order to maximize prediction from the child's scores on items. However, Achenbach e.a. (1981) found that the overall misclassification rate of 17.4% using dicriminant analysis is almost identical to the misclassification rate of 17.6% achieved by using the 90th percentile of the total behavior problem score on the Child Behavior Checklist. From this study it could be concluded that discriminant analysis only slightly improved the discriminative power of the instrument.

Another multivariate statistical technique widely used to detect relationships among variables to obtain a more refined measure of possible deviance is factor analysis. As this technique plays an important role in child psychiatric epidemiology, the next paragraph will be devoted to this

subject.

Compared with the multivariate statistical techniques, a simpler statistical approach to weigh scores is the one used by Sheperd e.a. (1971) and Kastrup (1976, 1977). These authors arbitrarily decided that any behaviors that occurred in fewer than 10% of the children were to be considered as deviant. Each child received a total deviance score by summing the scores for each item recorded on the questionnaire which was deviant on the basis of the 10% criterion.

II.2.4.3 Factor analytic approach

At the level of symptom groupings or syndromes, factor analysis has been used in a number of studies (Achenbach e.a., 1978; Langner e.a., 1976; Quay, 1980). Factor analysis is a multivariate statistical approach in which items are grouped together that intercorrelate highly with one another. In this way, mostly by applying checklists with a number of specific items to large representative clinical and non-clinical populations, empirically derived syndromes are obtained. In their review, Achenbach e.a (1978) compared empirically derived syndromes from diverse samples studied and found that 14 narrow-band syndromes and 4 broad-band syndromes had counterparts in a number of different studies. Empirically derived syndromes were also compared with a priori syndromes widely accepted in clinical use. Achenbach e.a. (1980 b) found a substantial but incomplete overlap between empirically derived syndromes and clinical syndromes listed in the Diagnostic and Statistical Manual III (DSM III, American Psychiatric Association, 1980). The still existing gap between the empirical and clinical approach can be regarded as a motive for refining both approaches.

The multivariate statistical approach has advantages but also disadvantages. Its results are dependent on the statistical technique chosen, the scoring of data, the number and content of the items, and the number, sex and age of the subjects used for the analysis. Therefore they are liable to variation. However, experience with the technique and its application is expanding. Comparison of existing data with data on different samples and the external validation of empirically derived syndromes will eventually have to provide further evidence of the usefulness of this approach. The strong advantage of the empirical approach is that it facilitates qualitative judgements about a child's behavior in a quick and relatively inexpensive way. For the study of large populations as in epidemiological research, the advantages are obvious. Furthermore, factor analysis renders it possible to make judgements on a more "molecular" level. For instance, a child not scoring in the pathological range when summing up all behavior problem items could score in the pathological range of a specific factor, such as hyperactivity.

Achenbach (1978, 1979), Achenbach e.a. (1979, 1983) and Edelbrock e.a. (1984) have linked this research approach with the practical use of empirically derived syndromes. The authors constructed the Child Behavior Profile, a standardized profile for portraying and categorizing behavioral

disorders and competencies. Taking age and gender into account, they factoranalyzed large samples of checklists filled out by parents (Child Behavior Checklist) or teachers (Teacher Report Form) of children referred to mental health settings. Behavior problem scales were then constructed using the factors derived through factor analysis and the scales were normed by obtaining checklists from parents and teachers of randomly selected children who had not received mental health services in the previous year. As outlined by Achenbach e.a. (1983), the Child Behavior Profile is widely utilized, mainly in the United States, not only for research purposes but also by mental health settings, as an aid to diagnosis and for purposes of accountability and planning. In epidemiological research, the Child Behavior Profile could be extremely valuable for case registers: systems of uniform data collection on all people referred for services within a delimited area. Because the Child Behavior Checklist, from which the Profile is scored, can easily be administered and is independent of the often diverse theoretical background of mental health settings, it is useful for administrative epidemiological purposes.

II.2.4.4 Clinical, diagnostic approach

In the absence of any objective criterion for defining child psychiatric disorder, judgement by an experienced psychiatrist is employed in some studies as the ultimate method of defining disorder. The so-called clinical, diagnostic approach to defining psychiatric disorder was adopted by Rutter e.a. (1970) in the Isle of Wight study. Psychiatric disorder was defined as "an abnormality of behaviour, emotions or relationships which was continuing up to the time of assessment and was sufficiently marked and sufficiently prolonged to cause handicap to the child himself and/or distress or disturbance in the family or community". The term "handicap" was defined as any disability which interfered in some way in the child's daily life. The judgement of the child's current functioning in relation to what is normal for his or her age, of the degree of handicap, and of the persistence of the behavior, is based on the clinician's knowledge, experience and skill. Although in the Isle of Wight study the interrater reliability for cases rated independently by two psychiatrists using a 4 point overall impairment scale was high (r=.89), it should be realised that the pychiatrists making the judgements came from the same clinic. It is unclear whether judgements made by psychiatrists having different clinical or theoretical backgrounds would show comparable reliability.

II.3 Sampling

Sampling, i.e. the selection of children to be studied, is the cornerstone of epidemiology. Four major sampling issues will be dealt with in this paragraph.

II.3.1 Representativeness

The importance of the results of an epidemiologic study is dependent on the degree to which the findings from a sample can be generalized to the whole population under study. Therefore, the sample must be representative of this population. This implies that every child in the population must have an equal opportunity of being selected as a subject of the sample. Some studies relied on "public school" listings as a sampling frame (Kellam e.a., 1975; Leslie, 1979; Shepherd e.a., 1971; Rutter e.a., 1970; Werry e.a., 1971), excluding children attending private schools and children who did not attend any school or who were institutionalized. Other studies used birth registers, census tract data or other listings covering a whole population either for random sampling (Achenbach e.a, 1981; Langner e.a., 1976; Richman e.a., 1982) or for selecting whole populations (Earls, 1980; Pringle e.a., 1966; Werner e.a., 1971). Investigators working in areas where birthregisters are present have the advantage of being able to cover accurately a total population of children.

II.3.2 Sample size

The determination of the sample size depends on the degree of precision required and on the estimated frequency of the variable of interest. For instance, in order to study child psychiatric disorders which have a low prevalence rate, such as psychotic disorders, one needs a much larger sample than for the study of overall rates of disorder. The sample should be sufficiently large in order to carry out a complete analysis. However, time, money and availability of staff are often limiting factors. Most studies of overall prevalence rates of child psychiatric disorders had sufficiently large samples considering the fact that the lowest estimated overall prevalence rate reported was 5.6% (Gould e.a., 1981).

II.3.3 Response rate

Because in every survey a certain number of informants will refuse to cooperate, a crucial question is whether the non-responders differ from the rest of the population. The fact that the non-responders do not differ from the main sample in age and social class can be misleading, if this leads to the conclusion that they do not differ in other respects too (Rutter, 1977). Cox e.a. (1977) found that non-responders differed from the main sample in that they tended to include a higher proportion of cases with behavioral deviance, reading backwardness, child or adult psychiatric disorder and marital discord. Therefore, it is important to cut down the number of non-responders as far as possible and to obtain as much relevant information as possible about those who do not cooperate.

II.3.4 Two-stage sampling

The most straightforward approach towards obtaining information from the general population is to investigate all members or the members of a sample which is representative of the population under study. However, if the aim of an investigation is to examine factors associated with psychiatric dysfunction in the general population by interviewing intensively parent and/or child, a two-stage sampling procedure has many advantages. This procedure, introduced in child psychiatric epidemiology by Rutter e.a. (1966; 1970; 1975), consists of screening procedures, mostly self-administered questionnaires applied to the total population in the first stage. The second stage involves more intensive investigation of children identified in the preceding stage. However, with every screening procedure employed in child psychiatric epidemiology a substantial number of children exhibiting mental health problems may be missed. Therefore, it is also important in the second stage to interview a sample of persons not selected as disordered by the screening procedure in order to estimate the proportion of children missed. Using screening instruments of known reliability and validity (see below), the two-stage sampling procedure is an economic way of obtaining information on large population samples. The two-stage sampling procedure was one of the procedures adopted for the study described here

II.4 Informant

Whereas in most adult psychiatric studies the person under study is the source of information, this is not the case in most child psychiatric studies. Most children, especially in the younger age groups, are considered unable to give reliable accounts on their behavior or emotions. Therefore, information about the child's mental health comes from other informants, especially the parents and teacher of the child.

Rutter e.a. (1968) investigated the agreement between ratings based on a psychiatric interview with the child with ratings based on a full study including parent and teacher reports. From a group of 31 severely disordered children, aged 7 to 12 years, 58.1% were judged having a marked, 32.2% a slight and 9.7% no abnormality on the basis of the child's interview. Although these data led Rutter e.a. (1968) to the conclusion that a psychiatric interview with the child without any background information can be used to make a provisional judgement on whether or not a child exhibits any psychiatric disorder, the authors found that the child's interview added little to the information already available from parents and teachers in the case of 7-to 12-year-old children. However, they also found that an interview with the child was essential in the case of 14-15-year-olds (Rutter e.a., 1976). Furthermore, for scholastic attainment tests and for neurological examination, the child as an information source is of course indispensable. Although it might seem self-evident to include information from the direct assessment of the child, the cost and effort of such assessment has to be weighed against the gain of information obtained

by it. Depending on the age of the target groups and aim of the investigation the loss of information by omission of the direct interview with the child can be negligible. However, this does not mean that an interview with the child is not highly important to reach an understanding of the meaning of a child's disorder or for making a choice of what kind of intervention is best for a particular child.

After discussing the child as an informant, two major sources of information in child psychiatric epidemiology will be discussed next: the parents and the teacher.

The parents in general are usually the people who know most about their child. They are informed about their child's behavior across many situations and across daytime as well as nighttime. Of the three intensive investigations performed in the Isle of Wight study (Rutter e.a., 1970), parental interview, direct psychiatric assessment of the child, and information from the teacher, parental interviews were found to be the most valuable contribution to the final diagnosis. Parental reports of the child's behavior will be influenced by the parents' attitude to the child's behavior, or their knowledge about their child's behavior especially outside the home. Parental hostility or denial of difficulties are major sources of bias. However, as Achenbach e.a. (1983) state, parents' views of their children's behavior are usually crucial in determining what will be done about it. Furthermore, problems like parental hostility arising in interactions with parents are probably important for children's long-term adaptation.

Teachers were found able to give accurate information on major aspects of the child's mental functioning (Rutter, 1977). However, this holds true only for teachers in primary school, because most of them see their children during a good deal of the day, whereas in secondary school there is a much more limited interaction with the child due to subject related teaching. In order to use teachers as informants, it is necessary that the investigation takes place at a time of the year when teachers are well acquainted with the child. As indicated in paragraph II.2.2 some behavioral problems not evident to parents may cause concern to teachers because these problems hamper social development and academic progress. Rutter e.a. (1970) found that there was a tendency for the parental questionnaire to miss antisocial children identified correctly by the teacher questionnaire. However, teacher reports also are susceptible to biases. Steinhausen e.a. (1983) found that teachers' evaluations of behavior disorders were dependent on socio-economic status of the child to a greater extent than parental reports.

In a review of studies providing prevalence estimates of childhood psychiatric disorders, Gould e.a. (1981) found that studies using primarily teachers as informants did not differ markedly in the average prevalence rates reported from studies using information from both parents and teachers. However, Gould e.a. (1981) stressed that this similarity in prevalence rates may be misleading because parent and teacher questionnaires applied to the same sample tend to select different children. Rutter e.a. (1970), as well as others (McGee e.a., 1984; Mitchel

e.a., 1966), reported a very small overlap (ranging from 7 to 20 percent) between groups of children selected on the basis of information from parents and teachers. The possible reasons for this lack of association between information obtained from parents and teachers are threefold: the situation specificity of the child's behavior, observation bias, and lack of efficiency of the instrument used. Mitchell e.a. (1966) therefore suggested that for estimating the distribution of child psychiatric disorders in the general population, the information of both teachers and parents ought to be utilized.

Of course the question arises which informant is the best source for judging the presence or absence of a child psychiatric disorder. As described above, all observations - whether they are done by the child himself, by the parents, teacher or clinician - are susceptible to different kinds of biases. In adult psychiatry, most studies rely on clinicians' observations and on the adult's own report as the major sources of information. In child psychiatry the situation is different. Achenbach e.a. (1983) states that the question is not how to obtain totally unbiased data on child behavior, but how to obtain the most useful data despite possible biases. There is not one single source of information we can call best in studying child behavior problems. Factors like the child's age and verbal ability but also factors like time, duration, place and circumstances of observations influence the observations as well as the attitude of the observer. One solution is to obtain information about a child from different sources such as parents, teachers and the child himself. However, a difficulty as yet unsolved is how to integrate this information coming in from different angles. In practice this problem is solved by making an a priori choice. Rutter e.a. (1970), for instance, used clinicians' judgements as the standard against which the value of the information from teachers and parents was measured.

II.5 Instruments

In order to measure the existence and type of pychiatric disorder in children, a number of instruments are available. Because of the large number of children involved, instruments used in child psychiatric epidemiology must be able to give an easily obtainable, reliable, valid and standardized account of the child's behavior. In this paragraph types of instruments and the quality of measurement in terms of its reliability and validity will be dicussed.

II.5.1 Types of instruments

Instruments used in child psychiatric epidemiology can be classified into screening instruments and instruments used for a more intensive inquiry.

For screening large populations questionnaires or checklists undoubtedly

play a major role, because the informants themselves or research-assistants can provide the necessary information at a relatively low economic cost.

Whereas in the epidemiology of adult psychiatric problems self-report questionnaires such as the General Health Questionnaire (Goldberg e.a., 1980) are now widely used, child psychiatric screening instruments are primarily directed at the parents or teachers as informants (see paragraph II.2). Screening questionnaires already mentioned in this review are the BSQ (Behaviour Screening Questionnaire; Richman e.a., 1971), the CBCL (Child Behavior Checklist; Achenbach 1966, 1978). The children's behaviour questionnaire for completion by parents (Rutter e.a., 1970) and for completion by teachers (Rutter e.a., 1967), the TRF (Teacher's Report Form; Achenbach e.a., 1983) and the questionnaire used by Langner e.a. (1976). Other instruments available for epidemiologic purposes are the Revised Behavior Problem Checklist (Quay e.a., 1983) and the Conner's Parent Questionnaire (Conner, 1973). For the latter two instruments as well as for the CBCL, syndromes empirically derived by factor analysis are available.

For more intensive inquiry instruments for interviewing parents and for direct assessment of the child are available. Graham e.a. (1968) developed and tested a parental interview used in the Isle of Wight study (Rutter e.a., 1970). In a modified version, this interview was used in the follow-up study by Richman e.a. (1982). This version also included questions on family factors such as parent-child interaction and marital discord; social and other stress factors; mental and physical health of parents.

A number of instruments for direct assessment of the child are available (Overachsel e.a., 1980). Hodges e.a. (1982 a,b) recently developed a structured child interview and observation form because existing instruments showed major limitations. Previous interviews either provided too general guidelines, such as the Rutter e.a. (1968) interview with the child, or the list of questions was too lengthy and direct for establishing good rapport with children.

Another approach to direct assessment of the child is behavior observation by a trained observer. However, this type of observational assessment is particularly liable to error due to the obtrusiveness of the observation itself, especially when carried out in the child's home (O'Leary e.a., 1979).

II.5.2 Quality of measurement.

A good assessment instrument should satisfy two major conditions: adequate reliability and validity.

1) Reliability.

The repeatability of a measure can be influenced by errors attributable to observer bias, subject's behavior variation across time and place, and variation as a result of observer-subject interaction. Therefore, the reliability should be assessed for every measuring instrument.

The most common forms of reliability reported in the literature of child

psychiatric disorders are (O'Leary e.a., 1979):

a) interrater reliability defined as the likelihood that two scorers or observers looking at the same behavior or same test protocol will arrive at the same score, and
b) test-retest reliability defined as the likelihood of obtaining the same score on two occasions.

Reliability can be assessed and expressed in many ways. The most primitive approach is by computing the percentage of agreement. However, this can be very misleading because a certain amount of agreement is to be expected by chance. An index of agreement between judges or scorers for nominal scales (for instance clinical psychiatric diagnoses) correcting for chance agreement is Cohen's kappa (Cohen, 1960). As the best measure of interrater reliability for ordinal and interval scales (for instance a five-point scale for recording severity of psychiatric disorder), the intraclass correlation coefficient is recommended (Tinsley e.a., 1975). This coefficient not only reflects differences in rank ordering of the scores - such as the product-moment correlation does - but also differences in magnitude (Bartko, 1976). Bartko gives an example of fictive scores showing perfect reliability as expressed by product-moment correlation coefficients, whereas the absolute magnitude of the scores differs markedly.

Reviewing the literature of major contributions to child psychiatric epidemiological studies of the general population, it is rather surprising that some studies do not report any reliability measures (Kastrup, 1976; Leslie, 1979; Sheperd e.a., 1971; Werner e.a., 1971). Percentage of agreement across scorers was used in some studies (Lapouse e.a., 1958; Rutter e.a., 1968; Graham e.a., 1968). The latter two studies also reported correlation coefficients between two raters respectively on the child interview (r=.84) and on the parent interview (r=.43). As explained above the correlation coefficient does not reflect differences in absolute magnitude of the scores across raters. Therefore Richman e.a. (1971) used a t test of difference between scorers in addition to a correlation coefficient (r=.77). Achenbach e.a. (1983) reported intraclass correlation coefficients for the Child Behavior Checklist as well as correlation coefficients in addition to t tests as a measure for test-retest, interparent and inter-interviewer reliability. Werry e.a. (1971) and Earls (1980) sufficed to report the use of instruments with already previously tested reliability.

Of course, it is important that efforts should be made to improve the instruments' reliability. One way to attempt this is with more "molecular" and objective items relevant to psychopathology. However, in this respect, the finding of Rutter e.a. (1968) is important. These authors found that contrary to what one would expect, some of the most global and differential items - such as "relationship with examiner" - were at least as reliable as some of the more objective items such as "smiling" or "number of spontaneous remarks".

2) Validity

Validity - the accuracy of measurement - can be tested in several ways

described below.

a. Content validity reflects the extent to which the measure represents the content the investigator intended to measure. Most instruments used in child psychiatric epidemiology are intended to cover a wide range of childhood behavioral and emotional problems thought to be of clinical relevance. Therefore, the most universal criterion used for determining content validity of an instrument is the ability of its items to discriminate between clinic versus non-clinic attenders. Nearly all child psychiatric epidemiologic studies report on the content validity of the instruments used.
b. Concurrent validity reflects the relationship between different measures of the same variable. For example, scores on one parent questionnaire can be correlated with scores on other parent questionnaires.
c. Construct validity reflects the extent to which an instrument measures a theoretical construct, for instance general ability in the case of an intelligence test or psychiatric disorder in the case of an instrument comprising items on behavioral problems. The sum of scores on an instrument designed to measure psychopathology should show good ability to identify children with psychiatric disorder. An instrument's sensitivity and specificity is a way of expressing its validity.
d. Predictive validity reflects the extent to which an instrument allows a prediction to be made about prognostic aspects, such as response to therapy, scholastic achievement, or natural course of the disorder. For instance, Richman e.a. (1982) found that high scores on their Behaviour Screening Questionnaire reflecting behavior problems at 3 years predicted clinically significant difficulties five years later in 61% of the children studied.

II.6 Classification

Once it has been determined that a child exhibits a psychiatric disorder, studies differ in the extent to which the child's condition is related to specific clinical diagnostic categories. Most studies of the prevalence of child behavior problems either report individual symptoms or overall psychiatric functioning without relating these to specific diagnostic categories. Exceptions are the studies of Leslie (1974), Rutter e.a. (1970) and Rutter e.a. (1976), although even in these the categories were rather broad and not operationally defined.

The Rutter e.a. (1976) follow-up study is an example of the importance of differentiating among psychiatric conditions. The 10-11-year-old children of the Isle of Wight study (Rutter e.a., 1970) were followed up when they were respectively 14 and 15 years old. It was found that there was only a slight rise in the overall rate of psychiatric disorder across time. However, there were two marked differences in the distribution of disorders, namely a substantial rise in the rate of affective disorders as

well as in the number of cases with school refusal.

In order to obtain agreement between researchers or clinicians on diagnostic terms, a classification system comprising clearly defined and meaningful syndromes is needed. The newest version of the Diagnostic and Statistical Manual, the DSM III (American Psychiatric Association, 1980) attempts to provide a diagnostic classification system of a priori child and adult psychiatric syndromes based on explicit exclusion and inclusion criteria.

DSM III has a separate section on major child psychiatric disorders. This seems to be an important step forward, especially advantageous for community surveys. However, DSM-III has some less positive aspects, too. Although DSM III diagnostic criteria for some adult psychiatric disorders were based on previously developed research diagnostic criteria, there were no such already existing research diagnostic criteria for child psychiatric conditions. DSM III consists of five axes: psychiatric syndromes on axis I; developmental and personality disorders on axis II; somatic conditions on axis III; severity of psychosocial stressors on axis IV; and highest level of adaptive functioning past year on axis V. The interrater reliability figures showed a kappa of .52 on axis I, for childhood disorders even when specific diagnoses were counted as agreeing if they fell within one broad category. This axis I kappa of .52 is much lower than the kappa of .72 found for the axis I diagnoses for the adult section of DSM III. One reason for the rather low interrater reliability figures could be that, compared with adult psychiatric conditions, child psychiatric conditions are much harder to score in a categorical yes-or-no fashion, because children tend to differ from each other much more in the degree to which they exhibit certain symptomatology.

Another approach is the use of empirically derived syndromes, already briefly discussed in paragraph II.2.4.3. These make it possible to retain more information in terms of dimensions of behavior compared with the use of all-or-none categories. The problem of dichotomising or not is also present in diagnosing a somatic condition like hypertension. In this example, it is clear that by forcing subjects into two categories - hypertensive subjects and subjects with normal blood pressure - we lose a lot of meaningful information which is retained when using a dimensional blood pressure scale.

II.7 Types of studies

Epidemiologic studies can be distinguished from clinical and experimental laboratory studies. Clinical studies are concerned with relatively severely diseased individuals, whereas epidemiology is applicable to the complete disease-spectrum in human populations. The main characteristic of experimental studies is the manipulation of independent variables, whereas in most epidemiologic studies the relationship between variables and diseases is studied in natural unmanipulated conditions. Although epidemiology is commonly equated with the study of the prevalence

of diseases or psychological problems, other research strategies are possible in epidemiology.

Once a statistical association between a characteristic and a disease or psychological condition has been demonstrated, e.g. in clinic patients, three major strategies can be applied to test this association: cross-sectional, retrospective and prospective research designs. In cross-sectional studies - sometimes called prevalence studies (see later) - as well as in retrospective studies - sometimes called historical or case-control studies - comparisons are made between a group of persons who at the time of the study have the disease or disorder and a group of persons who do not. The characteristics being compared differ between cross-sectional and retrospective studies, in that in the former the characteristics are present at the time of study, whereas in the latter it is determined whether or not the characteristics were present in the past.

For instance, Richman e.a. (1982, p.59) found maternal psychiatric problems to be much more common in the distburbed than in the comparison group of 3-year-old children. The mothers of the problem children were much more critical, less warm, and more depressed and anxious than the mothers of non-problem children. This part of their study shows a characteristic shared by most cross-sectional studies, in that the direction of the relationship remains unknown. In other words, from these results it cannot be determined which statement is true: do depressed, angry women produce psychological problems in their children or, the reverse, do problematic children produce depression and anger in their mothers ?

Another example of unknown directionality in a cross-sectional study was the association of severe reading retardation with antisocial behavior in the Isle of Wight study (Rutter e.a., 1970). However, it was not clear whether antisocial behavior was a result of reading problems or whether reading problems were caused by antisocial behavior.

Causal inferences can be made much stronger with retrospective study designs. For instance, Links e.a. (1980) compared the occurrence of pre- and perinatal complications in autistic children with their occurrence in same-sexed siblings. The authors used birth and pregnancy hospital chart records as well as maternal histories as sources of information. It was found that the autistic children had more complications in pre- and postnatal periods.

Although causal inferences can be made from retrospective surveys, there is a disadvantage in the fact that the retrospective approach is more liable to error. Bias can result from inaccurate information caused by incomplete or non-uniformly acquired information. Distorted memory about past events is another source of error for instance, when we use historical data from parents about their children.

Prospective studies, sometimes called cohort, longitudinal, or follow-up studies usually allow the strongest inferences to be made, because they have the advantage of a position at the beginning of a hypothesized causal chain. In such studies two groups of individuals not having the disorder under study are selected. In one group the hypothesized etiologic factor is present, in the other this factor is absent. At follow-up the rates of the

disorder in both groups are compared.

Richman e.a. (1982) followed up their sample of 212 intensively interviewed 3-year-old children at ages 4 and 8 years. One of the initial aims of this study was to clarify the mechanisms whereby reading retardation and antisocial behavior come to be associated by the age of 10 or 11 years. This study is an interesting example of how a cross-sectional survey - in this case the Isle of Wight study (Rutter e.a., 1970) - can generate new questions to be answered by a longitudinal approach. Unfortunately in the case of the Richman e.a. (1982) study, despite the enormous effort and clear methodology, the nature of the association between reading backwardness and antisocial behavior could not be clarified. Even at the age of 3 there was a strong association between language delay and behavior problems.

Although longitudinal studies can potentially yield direct estimates of the risk of developing a disorder, they have disadvantages too. Costs are high and the longitudinal approach takes a long time before definitive results can be obtained. The studies are liable to sample attrition. Furthermore, in a developing research branch like child psychiatric epidemiology existing assessment instruments will be improved or replaced by other instruments within the period the study is carried out. This makes comparisons with earlier studies difficult or obsolete.

II.8 Prevalence studies (cross-sectional studies)

Population studies of the prevalence of child psychiatric disorders not only provide data on the distribution of child mental health problems and on factors associated with the occurrence of these problems, they also provide baseline data for other investigations and they generate questions or hypotheses for further research. In table II.1, studies are summarized that provide prevalence data of child mental health problems in representative general population samples. Emphasis is put on studies providing overall rates of behavioral and emotional problems. Prevalence studies of specific ,and mostly rare, syndromes such as infantile autism, Tourette's disorder or elective mutism are not involved.

Prevalence is defined as the number of cases existing in a population at a specified time. Point prevalence refers to the number of cases present at a specified moment of time, whereas period prevalence consists of the point prevalence at the beginning of a specified period of time plus all new cases that occur during that period (Lilienfeld, 1976). Because most studies of the prevalence of child psychiatric disorders take a certain time to be carried out, they give an estimate of the period prevalence. Prevalence should be distinguished from incidence, which is defined as the number of new cases occurring in the population during a specified period of time.

A number of prevalence studies listed in table II.1 have already been mentioned in the preceding paragraphs. Some of the prevalence studies suffer from limitations. The major shortcomings are:

1. The reliance on public school listings as sampling frame (Leslie, 1979; Sheperd e.a., 1971; Rutter e.a., 1970; Werry e.a., 1971).
2. The reliance on teacher reports only (Werry, 1971).
3. The lack of data reported on the reliabilities used in the study (Kastrup, 1976; Leslie, 1979; Sheperd e.a., 1971; Werner e.a., 1971).
4. The small sample size (Earls, 1980; Kastrup. 1976).
5. The small number of specific behavior items on which prevalence rates were obtained (all studies except Achenbach e.a., 1981 and Langner e.a., 1976).

Of the preschool studies, the one reported by Richman e.a. (1982) is the most comprehensive (see paragraph II.7). The preschool study reported by Earls (1980), is limited by the small sample size (N=100), the unrepresentativeness of the population studied, and the lack of a predefined morbidity criterion.

Of the middle childhood studies, the ones by Lapouse e.a. (1958) and Rutter e.a. (1970, 1974) are the most widely cited. The studies by Rutter e.a. are characterized by the use of a multistage, multimethod approach. A number of factors thought to be responsible for the difference in prevalence rates between the Isle of Wight (12%) and Inner London (25.4%) populations were investigated.

The study by Achenbach e.a. (1981) is characterized by its systematic assessment of the effects of demographic variables on the prevalence of behavior problems and by its comparison between large representative samples of normal and disturbed children.

The National Child Development Study in the United Kingdom obtained behavior problem rates for a cohort of 17,000 children born during a preselected week in 1958. The data were analyzed at age 7 for 10,963 of the children (Pringle e.a., 1966; Davie e.a., 1972) and 16 for 11,608 of the children (Fogelman, 1976). However, the criterion of psychiatric disorder was arbitrarily defined in the first report (Pringle e.a., 1966) and not defined at all in the second (Fogelman, 1976). Furthermore, the number of specific symptoms reported by these studies was limited (see table II.1).

Sheperd e.a. (1971) report prevalence rates of arbitrarily defined overall psychiatric disorder and 37 specific symptoms in a large sample (N=6,287) of children. However, the study was restricted to children attending local authority schools, excluding children from private or special schools.

Cullen e.a. (1966) obtained behavioral problem rates on 3440 Australian children aged 0-15 years. Except for the reported rates, there was not much statistical exploration of the data.

Two studies, one in Sudan (Cederblad, 1968) and the other in Hawaii (Werner e.a., 1971) reported data on children who were mostly of non-Western ethnic background.

The studies by Leslie (1974) and Graham e.a. (1973) provide prevalence rates of psychiatric disorder in 13-14- and 14-15-year-old children, respectively. Both studies employed a two-stage sampling procedure (see paragraph II.3.4).

Because studies differ greatly in their definition of psychiatric disorder and in methods used, the diverse rates are difficult to compare with each other. In chapter V and chapter VIII our results will be compared with those in similar studies.

Table I.1 Prevalence Studies: methods & results.

Authors, yr.of publication Study Location	Sample size	Sample age	Sample method	Two stage	Definition of Disorder	Assessment Method(s)	Prevalence of psychiatric disorder overall[1]	Prevalence of psychiatric disorder specific symptoms	Prevalence of psychiatric disorder specific syndromes	Associated factors investigated
Achenbach e.a. 1981 U.S.A., Eastern States	1442	4-16	randomly selected homes using census tract data	-	statistical	parent questionnaire (child behavior checklist)	n.a.[2]	118 symptoms reported	n.a.	+
Cederblad, 1968 Sudan	1719	3-15	whole population 3 villages	+	individual symptoms	extensive study parent questionnaire; intensive study of subsample interview with mother.	n.a.	20 symptoms reported	n.a.	+
Connell e.a. 1982 Queensland, Australia	366	10-11	random sample of schools	+	clinical diagnostic	parent and teacher questionnaire in first stage; parent and child interview in second stage	10%	n.g.	n.a.	-
Cullen e.a. 1966, a,b Australia	3440	0-15	random sample of families	-	only individual items and mean number of behavior problems per child	parent questionnaire	n.a.	57 symptoms reported	n.a.	+

Davie e.a.; 1972; Pringle e.a. 1966 U.K.	10,963	7	whole sample of birth cohort using census data. School records by the time children were 7.	-	individual items on parents reports and arbitrarily defined cutoff score on teacher' social adjustment guide.	parent questionnaires; Bristol social-adjustment guide for completion by teachers; educational and medical assessment.	14% maladjusted	23 symptoms reported	10 syndromes reported	+
Earls, 1980 Island Community Northeastern U.S.A.	100	3	total population; families register.	-	statistical	parent questionnaire (Behaviour Screening Questionnaire) and child interviews	depending on definition of disorder	17 symptoms reported	n.a.	+
Fogelman e.a. 1976 U.K.	11,692	16	whole sample of birth cohort using census data.	-	individual items; no definition of disorder	parent and teacher questionnare	n.a.	18 symptoms reported	n.a.	-
Kastrup 1976 Aarhus County, Denmark	183	5-6	random sample of general population using municipal administration and hospital administration	-	statistical	parent interview	15% poorly adjusted	38 symptoms including developmental disabilities reported	n.a.	+

Krupinski e.a. 1967 Heyfield, Australia	756 children 241 adolescents	children and adolescents	whole population household census	-	clinical diagnostic	home interviews	7.4% in children; 12.3% in male adolescents and 3.3% in female adolescents.	n.g.[3]	3 broad categories reported	-
Langner e.a. 1976 New York, U.S.A.	1034	6-18	random sample households in one area	-	statistical and clinical diagnostic	parent questionnaire; direct child interview with subsample	11.5%	n.g.	empirically derived factors	-
Lapouse e.a. 1958, 1959 Buffalo City, U.S.A.	482	6-12	randomly selected homes using city directories	-	statistical	parent questionnaire and child interviews	n.a.	16 symptoms reported	n.a.	+
Leslie, 1974 Blackburn, U.K.	1198	13-14	total population of children attending local authority secondary schools	+	clinical diagnostic	parent questionnaire in the first stage; child interviews and teacher questionnaire in second stage	17.2% (21% in boys and 14% in girls)	n.g.	9 syndromes reported	+
McGee e.a. 1984 Dunedin, New Zealand	951	7	part of longitudinal study. Total sample of children born in one hospital.	-	statistical	parent and teacher questionnaire; behavior rating during psychometric assessment	30% high level of behavior problems	n.g.	3 broad categories reported	+

Miller e.a. 1974 Newcastle upon Tyne, U.K.	2.615	10	whole sample birth cohort using household census	-	disturbance reported by parents or recorded observations by members of the study team	parent interview and information from school	19.4% maladjusted	40 symptoms assessed, 5 reported	n.a.	+
Richman e.a. 1975, 1982 London borough	705	3	1 in 4 random sample of whole population using family register, excluding immigrants	+	clinical diagnostic	parent questionnaire (Behaviour Screening Questionnaire) in first stage; parent interview and language and developmental assessment of child in second stage.	7.3% At follow-up at 8 years of age, 61% of problematic 3-year-olds were still showing difficulties	24 symptoms reported	n.a.	+
Rutter e.a. 1970 Isle of Wight, U.K.	3316	10-11	total population of children attending public schools and children in certain administrative groups	+	clinical diagnostic	parent & teacher questionnaire in first stage. Individual child assessment; parent interview and information from teacher in second stage	6.8% excluding uncomplicated mental retardation and monosymptomatic disorders	31 symptoms reported	16 syndromes reported	+
Rutter e.a. 1974, 1975 Inner London borough	1689	10	total population of children attending public school	+	clinical diagnostic	teacher questionnaire in first stage. Parent interview in second stage	25.4%	n.g.	3 broad categories	+

Graham e.a. 1973 Rutter e.a. 1976 Isle of Wight, U.K.	2303	14-15	whole population	+	clinical diagnostic	parent and teacher questionnaire in first stage. Assessment of child; parent and teacher interview in second stage	13.2% in boys 12.5% in girls	8 symptoms reported	5 syndromes reported	+
Sheperd e.a. 1971 Buckinghamshire, U.K.	6,287	5-15	one-in-ten sample children attending local authority schools	-	statistical	parent and teacher questionnaire	39.3% boys and 42.1% girls without deviant behavior	37 symptoms reported	n.a.	+
Werner e.a. 1971 Kauai, Hawaii, U.S.A.	750	10	household census; identification of pregnant women; children followed up	-	interference with school achievement and individual items	home interview; teacher questionnaire	26.4% had one or more emotional problems; one out of every six had problems interfering with school achievement	5 symptoms reported	n.a.	+
Werry e.a. 1971 Illinois, U.S.A.	1753	5-8	school population	-	individual symptoms	teacher questionnaire	n.a.	55 symptoms reported	n.a.	-

1 overall = overall psychic functioning

2 n.a.= not assessed

3 n.g.= not given

CHAPTER III
POPULATION AND METHODS OF THE FIRST STAGE

III.1 Population

III.1.1 Background

The Netherlands is a delta country measuring about 41,000 sq.km. In 1981 the total population numbered 14.2 million. The country is divided into 11 provinces with population densities ranging from 159 to 1069 per sq.km (CBS, 1981). In 1977 the life expectancy of newborn boys was 72 years, that of newborn girls 78 years. Infant mortality has been brought down to 8.5 per thousand (van der Meeren e.a. 1981). Of the 11 provinces, the province of Zuid-Holland, containing the target population of this study, is the most densely populated one with its over 3,000,000 inhabitants. In this province, cities like Rotterdam, The Hague and Leyden are located. However, the southern and middle part of the province consists of rural and semi-rural areas. The total population of 4-16-year old children with the Dutch nationality living in the province of Zuid-Holland was 572,238 (51% boys; 49% girls) in 1981.

Although in 1981 there were 473,422 people not having the Dutch nationality who lived in the Netherlands (mostly Turkish and Moroccan), only those children who had the Dutch nationality were involved in this study. Foreign children were excluded from this study because language difficulties would create problems in obtaining reliable interview data. Furthermore, expected cultural differences, especially differences in the parents' toleration and judgement of their children's behavior, would be a factor complicating the comparison with Dutch children.

Infant, pre-school, and school health facilities are free for every child in the country. About 98% of pre-school children, 100% of primary school children and 90% of secondary school children receive school medical care (van der Meeren e.a., 1981). Once a mental health problem is detected by the general health facilities, the general practitioner or, the parents,

the child can be referred to the following specialized outpatient circuits: services for mentally handicapped children, pediatric departments of general hospitals, and child mental health services. National health insurance, compulsory for the middle and low income groups, makes these facilities financially available for virtually every child in need of specialized assessment or help.

As outlined in paragraph II.2.1 the developmental perspective is crucial when considering child psychopathology. A broad age range (4-16 years) was chosen in order to identify age-effects on the rates of child behavior problems and competencies. Below the age of 4, parental reports on behavior problems are less valuable due to variations in maturational rate and because of other idiosyncrasies caused by the young child's life environment which is much more restricted compared with the older child's much broader and less variable school- and social environment. Above age 16, the weakening of the emotional ties with the parents and the adolescents' moving out of the parental home, makes parents less valuable informants.

In this first (extensive) stage of the study, parental reports were chosen as the main source of information for reasons described in paragraph II.4. Teacher reports were also obtained for children aged 4-12 of the general population. This narrower age range was chosen because it was expected that secondary school teachers were less well informed about the child's behavior and hence would be less valuable. In this first stage of the study emphasis is placed on the parental report.

III.1.2 The selection of the general population sample

Using municipal birth registers 2600 4-16-year old children were selected using a stratified sampling procedure (100 children in each sex- and age-group) in two phases. The total number (572,238) of 4-16-year-olds living in the province of Zuid-Holland in 1981 led to a sampling fraction of 1:220.09. If this sampling fraction were used, interviewers would have needed to travel to many of the smaller municipalities in order to interview only one or two children. Therefore, a sampling interval of 2861 was chosen in order to obtain minimally 13 children (one in each age group with gender alternating) from the smaller municipalities. For each of the alphabetically arranged municipalities, the number of 4-16-year-olds could be computed (Centraal Bureau voor de Statistiek, 1981) and summed cumulatively. It was assumed that the alphabetic order would not affect the equal chance of every 4-16-year old child to be selected.

In the first stage, 86 of a total of 144 municipalities in the province of Zuid-Holland were selected when the number 2861 or a multiple of 2861 fell within the selected municipality's 4-16-year-old population. In the second stage the municipal administration of every selected municipality was asked to draw randomly a specified number of children with the Dutch nationality for whom the age and sex were indicated. Every child has a personal card in the municipal register. Sampling instructions were given

in order to obtain a random sample.

Because it was planned to conduct the interviews from February through May 1983, age groups were composed by choosing the mean of the interview period, namely April 1 as the birthdate at which a new age-group started. For instance the 4-year-old group was composed of children born from April 1 1978 till April 1 1979. At the time of interviewing a small proportion of children were expected not to have the same age (e.g. younger or older) as the age-group for which they were selected. By choosing April 1 this deviation was expected to be equal for both sides of the age spectrum.

The two largest cities, Rotterdam and The Hague, have computer-stored birth registers making the sampling procedure relatively easy and reliable. Of the 86 selected municipalities, 2 refused to cooperate, excluding 78 children from selection. One municipality first contacted the parents of the selected children to ask permission to give the child's name and address to our department. Five parents refused to cooperate.

In February, 1983, a letter was sent to the parents of the remaining 2517 children explaining the purpose of the study and indicating that an interviewer would come to the home to gather information on skills, hobbies, behavior and somatic complaints of the selected child.

III.1.3 The interviewing of the general population sample

From the end of February through May 1983 the interviews were conducted by 10 female and 4 male interviewers, previously instructed and trained. When contact with the parent was made, the interviewer presented a letter* referring to the previous one sent and summarizing the purpose of the study. The letter also indicated the interviewer's name, the name and address of the Sophia Children's Hospital as well as the assurance that all data would be strictly confidential. When the parent consented to the interview, the interviewer handed over the Child Behavior Checklist (CBCL) explaining that the CBCL contains questions on the child's skills, hobbies and school performance as well as questions on behavior. The interviewer first asked whether the child had been referred to a child mental health agency during the past year. The interviewer then asked each question on the CBCL and recorded the answer on a computer optical reader form. When in company with others, some parents preferred to give the answers without the questions being read aloud. In order to avoid undesirable bias, interviewers were instructed not to give examples or interpretations of behavior. If necessary, questions of parents were answered in order to help them describe their child's behavior. The interviewer was especially attentive to the parents' correct understanding of the scoring instructions and to their filling in the open-ended questions, if relevant. After the CBCL had been completed, the interviewer asked more detailed questions about a number of issues not reported in this study. At the end of the interview the parents were asked to sign a letter of consent indicating that they did not object to the teacher giving information on their child. Teachers were sent a questionnaire (TRF) in the same format as the CBCL.

* Appendix I

When parents were not at home, at least one visit was made after working hours thus avoiding a systematic drop out of children whose parents are both working at daytime. After a minimum of two subsequent failures to contact the parents at home, interviewers tried to contact the parents by telephone. The 71 parents who could not be reached even in this way were sent a CBCL with scoring instructions. In case parents had questions, the letter also indicated the name and telephone number of the person familiar with providing instructions to parents who fill in the CBCL as part of the intake procedure when their child is referred to our department.

In order to get information from the teachers of children in the general population sample, the Teacher Report Form (TRF, Achenbach e.a., 1983) together with the parents' letter of consent, were sent to the teachers of children 4-12 years. The first questionnaires were sent in April 1983. At the end of June, just prior to the summer holidays, a reminder was sent. Information was obtained at a time of the year when the teacher had known the child for a minimum of eight months.*

III.1.4 The general population sample

Of the 2447 parents reached, 2076 (84.8%) of the interviews were completed and could be used for data analysis. Of the main sample of 2600, the response rate was 79.8%. The remaining 20.2% included 78 children on whom municipal authorities refused to give information, 357 refusals by parents who explicitly refused or who did not respond to the CBCL sent by mail, 8 children on whom the interview could not be carried out because of language difficulties, 6 children who lived in a foster family or residential setting refusing cooperation, while the remaining 75 were untraceable (e.g. moved, house was empty, or wrong address). Of the 2076 children, 43 (2.1%) had been referred to a child mental health agency within the past 12 months. For some statistical analyses a normative sample was composed by excluding the 43 referred children from the general population sample.

The majority of respondents were mothers (89%). In 9% of the cases the respondent was the father, and in 2% "other" (e.g. foster parents, staff of residential setting). The ethnic background of the parents was 97% Caucasean, 2% Surinam, 1% Dutch Antilles, Turkish, Moroccan and other. The children all had the Dutch nationality.

Table III.1 shows the response rates according to age and sex of the children on whom CBCL's and TRF's were filled in. The N was obtained by subtracting from the original 100 in each age/sex group the following children: children on whom municipal authorities refused to give information, children on whom the CBCL could not be filled out because of language difficulties, and children who were untraceable. The overall response rate is 85.1% for the CBCL.

*Appendix II

Table III.1

Response rate in percentage, by age and sex of children in the general population sample for whom CBCL's and TRF's were filled in.

Age* group (yrs)	Boys N	Girls N	CBCL Boys (%)	CBCL Girls (%)	TRF Boys (%)	TRF Girls (%)
4	91	95	91.2	86.3	69.2	76.8
5	91	96	85.7	94.8	72.5	80.2
6	93	96	89.2	85.4	67.7	76.0
7	92	95	79.3	88.4	63.0	76.8
8	94	93	96.8	90.3	83.0	73.1
9	96	94	82.3	84.0	61.5	69.1
10	95	93	82.1	91.4	66.3	74.2
11	93	91	81.7	91.2	76.3	82.4
12	93	94	82.8	80.9		
13	95	96	81.1	86.5		
14	91	95	73.6	87.4		
15	94	95	80.9	74.7		
16	95	93	77.9	87.1		
Total	1213	1226	83.4	86.8	69.9	76.1

*For some statistical analyses actual age was used, excluding 5 children who were 3 years old at the date of the interview and 11 17-year-olds. These children are included respectively in the 4-year age group and 16-year age group presented in this table.

Socioeconomic status (SES) was scored on a six-step scale of occupation (ITS, 1975), as reported by the parent. If both parents worked, the higher-status occupation was used to score SES. Table III.2 shows the percent of each occupational level for the parents of children in the general population sample as well as for the normal and clinical sample (see paragraph III.5). The distribution of occupational levels in our samples is compared with that in a sample of males 35-44 years of age, living in the town of Zoetermeer, a commuter town in Zuid-Holland

(Epidemiologisch Preventief Onderzoek Zoetermeer, 1979). The age group of 35-44 years was chosen, because it corresponds with the age of the majority of fathers in our samples. Occupational levels were closely comparable between the two studies except for a category "shiftwork" in the Zoetermeer-study. The 6% falling within this category were equally distributed among occupational levels 1, 2, and 3.

Table III.2

Distribution of occupational level of parents of children in the general population sample, normal sample, clinical sample and of males in a comparison sample

Occupational level	general population sample N=2076 %	Normal sample N=2033 %	Clinical sample N=1387 %	Comparison sample N=406 %
1. Unskilled employees	6.2	6.3	4.1	5
2. Skilled manual employees	27.9	27.7	28.8	24
3. Clerical, technicians, minor professionals	20.5	20.3	27.6	34
4. Owners of small businesses	11.6	11.5	9.0	4
5. Supervisory, lesser professionals	18.1	18.2	15.9	18
6. Executives, major professionals, owners of large businesses	15.7	15.9	14.7	18

Table III.3 gives the distribution of urbanization for the general population sample as well as for the total population of the province of Zuid-Holland (CBS, 1981). As can be seen, the sample's distribution of urbanization is fairly consistent with the total population's distribution.

Table III.3

Distribution of urbanization for the total population of the province of Zuid-Holland, and the general population sample, and for children whose parents refused to cooperate

Urbanization	Zuid Holland	General population sample N=2522	Refusers N=363
	%	%	%
1. Rural; more than 20% of the population has agricultural profession	8.2	8.8	6.6
2. Semi-rural, less than 20% of the population has agricultural profession, fewer than 30,000 inhabitants	9.5	11.4	5.8
3. Suburban, less than 20% of the population has agricultural profession, more than 30% commuters	26.4	27.5	25.9
4. Small towns with historical character	0.1	1.6	1.1
5. Towns	54.5	50.7	60.6

★ Original sample excluding 78 refusals by municipal authorities.

The distribution of urbanization for those children whose parents or guardians refused to fill in the CBCL is also reported in table III.3. As can be seen, the proportion of towns is relatively higher in the sample of children whose parents refused to cooperate than in the general population sample. This difference is significant at a $P<0.001$ level (chi square=19.64, df=4). A cross tabulation of SES by urbanization showed a significantly higher proportion for the towns of lower SES children as measured by the father's occupational level (chi square=21.24, df=4, $P<0.001$). Hence, it can be inferred that there is a slight overrepresentation of lower SES children among the refusers.

This inference was supported by the following analysis. For every district in the city of Rotterdam a so called socio economic "backward score" is known (Das e.a., 1980). This "backward score" has been computed

from the number of unemployed, year of construction of the houses, number of people getting payment from endownment fund, and number of 17- and 18-year-olds receiving daytime education. Of the parents of 363 Rotterdam children reached, 86 were refusers. The distribution of socio economic backward scores in the responders- and refusers groups showed that there was a significant (chi square=6.42, df=2, $P<0.05$) overrepresentation of children living in socio economic backward districts among the refusers.

III.1.5 The clinical sample

As outlined in paragraph II.2, there are arguments in favor of considering referral status as good a morbidity criterion as any other available for child psychopathological conditions. Therefore, referral status is one of the morbidity criteria adopted for this study. Data on the clinical sample were obtained from 24 child mental health facilities, of which 20 were in the province of Zuid-Holland, whereas the other four were located in two other provinces. However, the mental health agencies outside the province of Zuid-Holland operated in urban and semirural environments comparable to Zuid-Holland. The settings were asked to have parents of newly referred children fill in the CBCL as part of their intake procedure. Comparable to the instructions given to the interviewers of the general population sample, the agencies were instructed to answer parents' questions. In this way the conditions under which the CBCLs were filled out were approximately the same.

Data from three agencies were excluded a priori from analysis because of expected bias due to patient selection. These agencies did not have every parent of newly referred children fill in the CBCL, but selected those parents who were expected not to get upset by the checklist. Contrary to our own experience with parents who fill in the CBCL as part of the intake procedure, these agencies judged the CBCL to be upsetting or undesirably intrusive for many parents.

The settings whose data were used included three university clinics, three clinics operated by independent hospitals, seven child guidance clinics, four community mental health centers, one private psychiatrist, and three day clinics for children aged up to six who are maladjusted or who show developmental problems.

Data collection was started in September 1982. At the end of October 1984, 1387 CBCL's could be used for data analysis. In this sample 56.6% of the respondents were mothers, 14.0% fathers, 22.4% both fathers and mothers and 7.0% "other". Ethnic background of the parents was 94.7% Caucasean, 1% Surinam and 4.3% other. The distribution of occupational level for the parents of children in the clinical sample is shown in table III.2. For the distribution of children by age and sex of the clinical sample on whom CBCL's were filled in, see table III.4.

Table III.4

Distribution by age and sex of children in the clinical sample for whom CBCL's were filled in.

Age group (yrs)	Boys	Girls	Age group (yrs)	Boys	Girls
4	53	25	11	84	54
5	73	23	12	89	41
6	63	30	13	84	42
7	75	43	14	63	45
8	90	46	15	39	33
9	95	32	16	13	14
10	98	40	total	919	468

III.1.6 Discussion

In an epidemiological study like this, the quality of the samples is of crucial importance. The procedure we followed with the selection of our general population sample gave every child aged 4-16 in the province of Zuid-Holland an equal chance to be selected. A check on the representativeness of the start population is its distribution of urbanization, which is closely similar to the distribution of the total Zuid-Holland population. When we consider that our sample was selected from the general population, the response rate of 85.1% for the CBCL is satisfactory.

Achenbach e.a. (1981) obtained a response rate of 82.3% for their general population sample. Rutter e.a. (1970) obtained a response rate of 88.5% of parents who completed the questionnaire. However, Rutter's sample was drawn from the public school population and missed children attending private schools. They found that a substantial number of parents who did not return the questionnaire had had difficulty in completing the forms unassisted. This kind of bias was removed from our sample because the parents were interviewed.

The distribution of occupational level in our general population sample was closely similar to the SES distribution in a sample of males aged 35-44 years in the town of Zoetermeer. The lower percentage of owners of small businesses and the higher percentage of clerical and minor professionals in the comparison sample may be due to the fact that Zoetermeer is a commuter

town with a large population increase during the last 20 years.

By comparing the distribution of the degrees of urbanization in the non-responder and responder samples in two independent ways, it was found that the proportion of lower SES children was higher in the non-responder group. Because lower SES children show more behavior problems and less social competence and school achievement, as is shown in this study, the overrepresentation of lower SES children among the responders may influence the data of the general population sample. However, this slight overrepresentation of lower SES children in a group of 363 refusers will exert a minimal influence on the data of the 2076 who did respond.

III.2 Methods

III.2.1 The Child Behavior Checklist (CBCL) and the Teacher Report Form (TRF)

The Child Behavior Checklist (CBCL) is a questionnaire developed by Achenbach (Achenbach, 1966; Achenbach e.a., 1983). It was designed to report in a standardized way the behavior of children aged 4 through 16 as assessed by their parents. The CBCL contains behavioral items which are relevant to researchers and practitioners in the field of developmental psychopathology. The CBCL consists of two parts: 20 social competence items and 118 behavior problem items (Appendix A). It was translated into Dutch with the help of a linguist in order to keep as closely as possible to the behavior covered by the original items.

The social competence items were designed to measure a child's positive behavioral characteristics with a minimum of social desirability effect in the parents' answers. Parents were asked to specify the number of sports, hobbies, participation in organizations, jobs and friendships the child likes to take part in, as well as to estimate the quality and the amount of involvement in each activity. In addition, parents were asked to specify how well their child gets along with siblings, other children, and parents and how well their child can play and work by himself/herself. Questions concerning academic performance were: current school performance, special class attendance, grade repitition and other problems in school.
The behavior problem section of the CBCL consists of 118 non-redundant items describing a broad range of problem behavior which is of concern to parents and clinicians. Some items required parents to describe the child's behavior, making it possible to correct the parents' scoring when the described behavior did not fit the item. The scoring instructions given by Achenbach e.a. (1983) were followed in this procedure. Every CBCL completed was checked for inappropriate scoring and was corrected if necessary. Two items (56, other physical problems without known medical cause; and 113, any problems the child has that are not listed in the checklist) made it possible for parents to add behavioral problems not elsewhere indicated. A 0-1-2 scale was used to score responses. The scored behavior describes the child within the past 6 months. A 0 is scored by the parents if the

description is not true for their child, a 1 if it is somewhat or sometimes true, and a 2 if it is very true or often true.

The CBCL can be filled in by parents or others who know the child well, in 15-30 minutes.

The Teacher Report Form (TRF) was also developed by Achenbach e.a. (1983) and was designed to describe behavior problems and school performance of children in a standardized way. It was translated into Dutch. The TRF, wich is filled in by the teacher, contains two sections: the first part consists of questions on the context in which the teacher knows the child, previous special services, repetition of grades, academic performance, and general adaptive characteristics of the child. The second part of the TRF contains behavior problem items in the same format as the CBCL. The teachers' ratings are based on the previous 2 months. Some items such as "nightmares" which are only scorable by parents are replaced by items more relevant to the school situation such as "overconforms to rules" or "inattentive, easily distracted". In this study only the behavior problems scores on the TRF are used for data analysis.

III.2.2 Reliability

In order to test the repeatability of the scores on the CBCL and TRF, intraclass correlation coefficients (Bartko, 1976) were computed to assess test-retest, inter-interviewer, and interparent reliability for the social competence items and the behavior problem items of the CBCL and to assess test-retest reliability for the behavior problem items of the TRF.

In order to assess test-retest reliability each interviewer asked for permission at the end of 10 consecutive interviews to return 3-4 weeks later to administer the CBCL again. The interval of 3-4 weeks was rather arbitrarily chosen because after a shorter interval there might be a greater chance that parents remember their answers given on the first occasion, whereas after a longer time interval the child's behavior might have changed. Parents for retest interviews were randomly chosen. For 117 of the 130 children retests were obtained.

Although it was expected that inter-interviewer reliability would be high due to the fact that no judgement about a child's behavior was allowed to be made by the interviewers, we wanted to be sure whether systematic scoring errors or some unknown factor might cause inaccuracies. Therefore for 58 randomly selected children pairs of interviewers independently scored the parents' answers on the CBCL on the same occasion.

To assess interparent agreement 23 fathers and mothers of children referred to our department independently filled in the CBCL at intake. The receptionist made sure that no discussion could have taken place between the parents. During a 5 month period all parents were chosen whose child was seen separately at intake, while they were both waiting in the waiting room. Only a minority of parents underwent this intake procedure, often concerning children who were seen in a consultation procedure or whose problems were of a kind not acceptable for being put on a waiting list.

For the TRF, a test-retest reliability was assessed by mailing 44 checklists at a 3-4 week interval to randomly selected teachers who completed the TRF on the first occasion on children from the general population. A reminder was sent to those teachers who did not return the TRF after the first request. Twenty-two TRF's were used for assessing test-retest reliability.

Intraclass correlation coefficients were computed from one-way analyses of variance (ANOVA) for total behavior problem scores obtained by summing the 0's, 1's or 2's across the 118 items on the CBCL as well as on the TRF. Intraclass correlation coefficients were also computed for the total social competence score. The total social competence score was computed in the same way as indicated by Achenbach e.a. (1981).

The formula given by Bartko (1976) for the one-way ANOVA intraclass correlation, which was used in this study is:

$$ICC = (MSB-MSW) / [MSB+(C-1)MSW]$$

ICC = intraclass correlation coefficient, MSB = mean square between subjects, MSW = mean square within subjects, C= number of raters. Table III.5 gives the ICC's for the reliabilities.

Table III.5

Intraclass correlation coefficient (ICC) for reliabilities of total scores on the CBCL and TRF

Reliability	N	ICC
CBCL-behavior problems		
Test-retest	117	0.75
Inter-interviewer	58	0.99
Inter-parent	23	0.70
CBCL-social competence*		
Test-retest	104	0.77
Inter-interviewer	55	0.99
Inter-parent	19	0.69
TRF-behavior problems		
Test-retest	22	0.84

*The number of cases in the analysis of the social competence scores is smaller because of missing data

III.2.3 Statistical methods used in the analyses

To assess the effects and interaction of age and gender on the prevalence in the general population of specific items on the CBCL, analyses of covariance (ANCOVA) were carried out with SES as a covariate (a 13 (age) X 2 (gender) factorial design with SES as covariate).

A clinic versus non-clinic comparison was made by performing ANCOVA on the combined normal sample (=general population sample without referred children) and clinical sample with age, gender and clinical status as main effects and SES as covariate (a 2 (clinic vs non-clinic) X 13 (age) X 2 (gender) factorial design with SES as a covariate).

In order to estimate the power of the CBCL to identify children resembling clinical and non-clinical groups of children, discriminant function analysis was employed for the normal and clinical samples. A random sample of 1387 children was drawn from the normal sample. Although both normal and clinical samples had equal numbers of children in this way, their age distributions were not the same due to the unequal distribution of the clinical sample.

In order to compare our results with those of Achenbach e.a. we grouped the children for each sex in the following age ranges: 4-5, 6-11, 12-16 years. These age ranges were chosen because they correspond with different types of education. It can be argued that the 6-11 and 12-16 years age periods are rather broad and contain children varying in biological and psychological developmental levels. However, breaking these age groups into smaller units introduces statistical problems due to the smaller number of subjects, especially when discriminant function analysis is employed or, as Achenbach did, factor analysis. The classification derived by discriminant analysis in half of the sample was cross-validated by applying the weights to the other half of the sample. All even numbered subjects were assigned to the sample on which the discriminant function was to be derived and all odd numbered subjects were assigned to the sample on which the classification was cross-validated. A stepwise discriminant function analysis was computed for each sample. The behavior problem items were entered in the order of their reduction of Wilks's lambda until items no longer contributed significantly to the reduction of Wilks's lambda ($P<0.01$). For some age-sex groups the number of items used in the analysis needed to be reduced first in order not to have the program being stopped due to the large size of the subjects-variables matrix. Items that did not have a significant ($P<0.01$) clinical status effect by ANOVA were excluded from the discriminant analysis in the 6-11-year-old boys and girls (table III.6). ANOVA was employed on the 6-11-year-old combined clinical and normal sample (N=1690).

For all statistical analyses performed in this study the Statistical Package for the Social Sciences (SPSS and SPSS-X) was used on a DEC-20 computer.

Table III.6

Items excluded from discriminant analysis because P value of clinical status effect by ANOVA was smaller than 0.01. Boys and Girls 6-11 years.

4 Asthma	35 Feels worthless
5 Behaves like opposite sex	77 Sleeps much
7 Bragging	98 Thumb-sucking
28 Eats nonfood	99 Too concerned with neatness
32 Needs to be perfect	105 Alcohol or drugs
34 Feels persecuted	110 Wishes to be opposite sex

III.2.4 Discussion

The intraclass correlation reliability coefficients we found for the behavioral problem section were nearly the same as those for the social competence section. Compared with Achenbach e.a.(1981) our reliability coefficients are lower. These authors found ICC's for test-retest with a one-week interval of .952 for behavior problems and 0.996 for social competence. The ICC's for interparent reliability were: 0.985 for behavior problems and .978 for social competence. Achenbach used a slightly different method for computing the ICC's, and the longer time interval between test and retest in our study (3-4 weeks) could have influenced our test-retest reliabilities negatively.

There are a number of differences between American and Dutch societies, which could have influenced the scoring. In the United States, the filling in of checklists could be much more a matter of course than in the Dutch society. Relative unfamiliarity with the procedure may have influenced the reliabilities in our study, although the general impression of the interviewers was that most parents readily understood the questions. Furthermore, as we shall see in the next chapters the similarities between the American and Dutch studies in mean total scores and frequency distributions are striking.

Our reliability data of behavior problem items on the CBCL are somewhat higher than the ones found by Rutter e.a. (1970) for their parent questionnaire. Using ordinary product-moment correlations, which is not recommended any more for the assessment of reliabilities, the authors found a test-retest reliability (r) of +0.74 and an interparent reliability (r) of 0.64. For the teacher questionnaire they found a test-retest reliability (r) of 0.89, compared to the intraclass correlation of 0.84 for test-retest reliability of the TRF in our study.

CHAPTER IV
RESULTS OF THE FIRST STAGE

IV.1 Behavior problems in the general population.

IV.1.1 The prevalence of specific behavior problems.

Figures 1-112 (Appendix B) show the percentages of children from the general population grouped by age and gender for whom each behavior problem listed in the CBCL was reported by the parents. The scores of 1 and 2 were combined to provide the percentages. In order to save space and to make our data comparable to Achenbach e.a.'s data, two ages were taken together to form 2-year intervals except for age 16, which is presented separately. It is emphasized that although the points in the figures 1-112 are connected for clarity of presentation, the results should be taken as they are, namely cross-sectional frequencies of behavior problems. This means that the findings do not automatically imply that the behavior of an individual child will follow the course as indicated in the figures. The percentages depicted in the figures were standardized for SES in tertiles. This was done to visualize the frequencies related to age and gender, while sample differences in SES were partialled out.

IV.1.2 Total behavior problem scores

The total behavior problem score is calculated by summing all 0's, 1's and 2's scored on the 118 items for each individual.

Appendix C presents the mean total behavior problem score for each age

and sex for the general population sample. The results show that the mean total behavior problem score decreases with age except that the 7-through 10-year-olds show a rise in the mean total behavior problem score.

Except for the 16-year-olds, boys have a slightly higher mean total score than girls. As will be shown in the next paragraph the gender and age effects are statistically significant ($P<0.001$) in ANCOVA.

In order to report the norms for the "healthy" (i.e. nonreferred) population, Appendix C also gives the mean total behavior problem scores by age and sex for the general population sample without the 43 referred cases.

IV.1.3 Analyses of covariance (ANCOVA).

Analyses of covariance (ANCOVA) were performed on the scores of each item and on the total behavior problem score of the general population sample (N=2076) in order to assess the main effects and interaction of age and gender, with SES as covariate. Appendix D shows the results for the main effects, gender and age, as well as for the effect of clinical status, as will be outlined in paragraph IV.3.1. When a P value of 0.01 is accepted, gender had a significant effect in 41 items and total behavior problem score. Age effect was present in 56 items and total behavior problem score, while SES was a significant source of variance in 35 items (see table IV.3), and total behavior problem score. All significant SES effects showed higher scores for lower SES. For the following 4 items, significant gender x age interactions were found: 37, Fighting; 72, Sets fire; 102, Underactive; and 105, Alcohol or drugs.

The probability of making a type I error (rejecting the null hypothesis when in fact it is true) increases as the number of independent significance tests increases. Some authors, such as Achenbach e.a. (1981) and Feild e.a. (1974), endorse the discarding of significant effects having the smallest F values. In our case 5 of the 119 effects in ANCOVAs could reach the 0.01 level of significance by chance. Therefore, the 5 smallest F values for each main effect were marked with a superscript in Appendix D.

As can be seen from Appendix D, the percentages of variance accounted for by gender and age effects are small. According to Cohen's criteria, effects accounting for 1%-5.9% are considered small, effects accounting for 5.9%-13.8% are considered medium, and effects accounting for more than 13.8% are considered large (Cohen, 1977).

For only one item (7, Bragging) the gender effect could be considered medium. For 17 items gender reached the criterion for small effects, while for the remaining 23 items and total behavior problem score, gender accounted for less than 1% of the variance.

For three items (29, Fears; 74, Showing off; and 98, Thumbsucking) age effects were medium, whereas for the remaining 53 items, plus total behavior problem score, the age effects were considered small. None of the significant age effects accounted for less than 1% of the variance.

The associations of our items for which gender, age or SES effects were found, with the two empirically derived broad band syndromes called

Externalizing and Internalizing by Achenbach (1978) and labeled by other authors Aggression versus Inhibition syndromes, are listed in tables IV.1-IV.3. These broad band syndromes were derived by factor analyses of CBCL behavior problem items reported by parents of large samples of clinically referred children (Achenbach, 1978; Achenbach and Edelbrock, 1979). We could take advantage of the fact that our CBCL items were the same as those used by Achenbach e.a. in their factor analyses.

Table IV.1
Behavior problems with gender differences ($P<0.01$) in ANCOVA's Association with Externalizing and Internalizing syndromes+

	Boys scoring higher than Girls		
Externalizing		Internalizing	Neither
1. Acts too young	43. Lying	36. Accident prone	6. Encopresis
7. Bragging	57. Attacks people	59. Plays with sex parts in public	15. Cruel to animals
8. Can't concentrate	62. Clumsy		38. Is teased
10. Hyperactive	74. Showing off		60. Plays with sex parts too much
16. Cruel to others	79. Speech problem		96. Sex preoccup.
20. Destroys own things	90. Swearing		108. Wets bed
21. Destroys others' things	94. Teases a lot		
22. Disobeys at home	95. Temper tantrums		
23. Disobeys at school	104. Loud		
37. Fights	105. Alcohol drugs		
41. Impulsive	106. Vandalism		
	Girls scoring higher than Boys		
Externalizing		Internalizing	Neither
5. Acts like opposite sex		29. Fears	4. Doesn't eat well
		56.F Stomach problems	55. Overweight
		75. Shy, timid	64. Prefers young kids
		88. Sulks	98. Thumbsucking
			99. Too neat
			110. Wishes to be opposite sex

+ Empirically derived broad band syndromes; source Achenbach 1966, 1978; Achenbach and Edelbrock 1979. Items assigned on the basis of highest factor loading reported in these studies.
Boys score higher on externalizing items, Girls higher on internalizing items, chi square = 13.52, df=1, $P<0.001$.

Table IV. 2

Behavior problem items with age differences (P<0.01) in ANCOVA's. Association with Externalizing and Internalizing syndromes.[+]

Younger children scoring higher

Externalizing	Internalizing	Neither
3. Argues a lot	11. Too dependent	6. Encopresis
7. Bragging	14. Cries a lot	24. Doesn't eat well
10. Hyperactive	29. Fears	36. Accident prone [G‡]
19. Demands attention	36. Accident prone [B‡]	58. Picking
20. Destroys own things	47. Nightmares	59. Plays with sex parts [G] in public
21. Destroys others' things	59. Plays with sex [B] parts in public	63. Prefers older children [B]
22. Disobedient at home	70. Sees things	83. Hoarding [G]
27. Easily jealous	83. Hoarding [B]	98. Thumbsucking
63. Prefers older children	88. Sulks	107. Wets self
68. Screams	109. Whining	108. Wets bed
74. Showing off		
79. Speech problems		
93. Talks much		
104. Loud		

Older children scoring higher

Externalizing	Internalizing	Neither
1. Acts too young	35. Feels worthless	44. Nailbiting [G]
39. Bad friends	42. Likes to be alone	53. Overeating
44. Nailbiting [B]	51. Dizzy	
61. Poor schoolwork	56. Headaches	
101. Truancy	102. Underactive	
105. Alcohol, drugs	103. Unhappy, sad	
	112. Worrying	

+ Empirically derived broad band syndromes; source Achenbach 1966, 1978; Achenbach and Edelbrock 1979. Items assigned on the basis of highest factorloading reported in these studies

‡ B=Boys, G=Girls

Table IV. 3

Association with Externalizing and Internalizing syndromes [+] of behavior problem items on which lower SES children score higher (P<0.01) in ANCOVA's.

Externalizing		Internalizing	Neither	
3. Argues a lot	38. Is teased [G‡]	9. Obsessions	26. Lacks guilt [B‡]	
5. Behaves like opposite sex	44. Bites fingernails	11. Too dependent	38. Is teased [B]	
8. Can't concentrate	45. Nervous [B]	13. Confused [B]	44. Bites fingernails	
10. Hyperactive	64. Prefers [B] young kids	14. Cries a lot	53. Overeating	
13. Confused [G]	68. Screams	45. Nervous [G]	58. Picking	
16. Cruel to others	86. Stubborn	65. Refuses to talk	64. Prefers young kids [G]	
19. Demands attention	90. Swearing	88. Sulks	96. Sex preoccup. [B]	
20. Destroys own things	93. Talks too much	92. Walks, talks in sleep [B]	99. Too neat [G]	
22. Disobedient at home	94. Teases	99. Too neat		
23. Disobedient at school	95. Temper tantrums	109. Whining		
26. Lacks guilt [G]	96. Sex preoccup. [G]			
37. Fighting	104. Loud			

+ Empirically derived broad band syndromes; source Achenbach 1966, 1978; Achenbach and Edelbrock 1979. Items assigned on the basis of highest factor loading reported in these studies.

‡ B=Boys, G=Girls

IV.2 Social competencies in the general population sample as reported by parents.

IV.2.1 Total social competence scores

The total social competence score is calculated in the same way as outlined by Achenbach e.a. (1981). Appendix E presents the mean total social competence scores by age and sex for the general population sample (N=2076), as well as for the "healthy" (i.e. nonreferred) population (N=2033). The sharp rise in social competence scores after age 5 is caused by the exclusion of the school scale for the 4- and 5-year-olds. After age 5 there is a slight rise in the mean scores, declining again after age 12. This age effect is not significant ($P>0.01$) in ANCOVA (see next paragraph). Girls generally score higher than boys. This difference is statistically significant ($P<0.001$) in ANCOVA (see next paragraph).

IV.2.2 Analyses of covariance (ANCOVA)

ANCOVAs were employed to assess the effects of gender and age on social competence scores in the general population sample. Appendix F summarizes the results for the main effects as well as for the effect of clinical status as will be reported in paragraph IV.3.1. The significance level was accepted at $P<0.01$.

For 7 items there was a significant ($P<0.01$) gender effect, 3 of which were small effects according to Cohen's criteria, while the remaining 4 accounted for less than 1% of the variance. For total activities scale, the total social scale and for the total social competence score the gender effects accounted for less than 1% of the variance. The gender effect for the total school scale was small according to Cohen's criteria. All gender effects were in favor of the girls, who scored higher on social competence.

For 14 items a significant age effect was found ($P<0.01$). According to Cohen's criteria age effects were medium for the following 6 items: I A, Number of sports; III A, Number of organizations; III B, Participation in organizations; IV A, Number of jobs; IV B Job performance; VII 3, (No) grade repetition. For all other items showing significant age effects as well as for total activities scale, total social scale, and total school scale, the age effects were small. There was no significant age effect for the total social competence score. Of the linear age effects of the activities section, only item III A (Number of nonsports activities, e.g. hobbies) was in favor of the younger children. The quality of the contacts with friends is reported to be in favor of the younger children, whereas the reverse is true for the quality of the behavior with parents. Academic performance scores are reported to be in favor of the younger children, which is also reflected in the total school scale. Total activities scale and total social scale show higher scores for older children.

All of the 10 significant ($P<0.01$) SES effects found were in favor of higher SES children. These 10 items are listed in table IV.4. No

significant SES effects were found for the total scales.

For the following 3 items, two-way interactions (P<0.01) between gender and age were found: IA, Number of sports; IIA Number of nonsports activities; and IIIA Number of organizations.

Table IV.4

Social competence items with SES effects (P<0.01) in ANCOVA's.

All effects were in favor of higher SES children

IA	Number of sports	IIIA	Number of organizations
IC	Skill in sports	IIIB	Participation in organizations
IIA	Number of nonsports activities	VII 1	Academic performance
IIB	Participation in activities	VII 2	(No) special class
IIC	Skill in activities	VII 3	(No) grade repetition

IV.3 Comparison between referred and nonreferred samples

IV.3.1 Analyses of covariance (ANCOVA)

The two groups of children, referred (N=1387) and nonreferred (N=2033), were compared in a number of ways. First, ANCOVAs were performed on the total behavior problem score and on the scores obtained on each CBCL-item with a 2 (referred vs. nonreferred) x 13 (age) x 2 (gender) factorial design and SES as covariate on the whole sample (N=3420). The normal sample consisted of the general population sample without the 43 children referred to a mental health agency past year. For the vast majority of items the effect of clinical status was highly significant (P<0.001). The results are shown in Appendix D. Of the 119 ANCOVAs performed, no significant differences (P<0.01) were obtained for the following four items: 5, Behaves like opposite sex; 98, Thumbsucking; 105, Alcohol or drugs; 110, Wishes to be of opposite sex. For all other items higher scores were obtained for the referred children. For the following 4 items, significance levels smaller than 0.01 were reached: 4, Asthma; 32, Needs to be perfect; 55, Overweight; and 99, Too concerned with neatness of cleanliness. All other items had P values smaller than 0.001. Although significant, the size of the effects differed in terms of percentage of variance. According to Cohen's criteria for 15 items the clinical status effects were large (accounting for more than 13.8% of the variance). For the sake of clarity these items are shown in table IV.5.

Table IV.5

CBCL behavior problem items for which clinical status showed large (more than 13.8% of the variance) effects in ANCOVAs

Item		Percentage of variance	Item		Percentage of variance
25	Poor peer relations	24.0	26	Lacks guilt	15.2
103	Unhappy, sad	22.1	61	Poor school work	15.2
45	Nervous	21.2	112	Worrying	15.2
69	Secretive	20.3	8	Can't concentrate	14.4
9	Obsessions	16.8	13	Confused	14.4
19	Demands attention	16.8	48	Not liked	14.4
34	Feels persecuted	16.0	84	Strange behavior	14.4
50	Too fearful, anxious	16.0			

Table IV.6

CBCL-items showing the smallest difference (less than 1% of the variance) between clinical and non-clinical groups in ANCOVAs

Item		Item	
2	Allergy	56[e]	Skinproblems
3	Asthma	77	Sleeps much
5	Behaves like opposite sex	83	Stores up unneeded things
28	Eats nonfood	92	Talks or walks in sleep
32	Needs to be perfect	98	Thumbsucking
42	Likes to be alone	99	Too concerned with neatness or cleanliness
44	Bites fingernails	105	Alcohol or drugs
53	Overeating	110	Wishes to be of opposite sex
55	Overweight		

For 37 items the effect was medium. For 49 the effect was small, whereas for the remaining 13 items showing significant effects, the effects accounted for less than 1% of the variance. The 17 items showing the least differentiation between clinical and non-clinical groups are listed in Table IV.6. The nonsignificant effects ($P>0.01$) are also included in this table.

For the total behavior problem score, clinical status accounted for 40% ($P<0.001$) of the variance. Although age had a significant effect on the total behavior problem score in the general population sample, the age effect was not significant ($P>0.01$) in the combined clinical and non-clinical sample. Gender accounted for 3% ($P<0.001$) of the variance (small effect) of the total behavior problem score in the combined sample, boys scoring higher than girls.

ANCOVAs were also performed on the social competence scores on the combined clinical and non-clinical sample. Appendix F shows the effect of clinical status, which is significant at a $P<0.001$ level for 18 of the 20 social competence items. For the following items clinical status did not reach a significant ($P<0.01$) level: II A, Number of nonsports activities; and II B, Participation in activities. Note that item IV A, Number of jobs and IV B, Job performance were the only items for which the clinical status effect indicated higher scores for the referred children. This is probably due to the fact that unfortunately we had not clearly instructed the interviewers of the normal sample that these questions concerned paying as well as nonpaying jobs and chores. Therefore a number of interviewers only scored paying jobs, whereas parents of referred children were instructed to include nonpaying jobs as well.

The size of the effects expressed in percentages of variance was largest for the social and school scales. Five of the 6 social items showed a medium effect. One school item (VII 4, (No) other school problems) showed a large effect, whereas two items (VII 1, Academic performance; VII 2, (No) special class) showed medium effects, and one ((No) grade repetition) a small effect. Of the 8 significant effects of the activities scale, 6 were small; for 2 items the effect accounted for less than 1% of the variance. For the total activities scale, the clinical status effect was nonsignificant. The largest effect of clinical status was on the school scale, even larger than the effect for the total social competence score.

IV.3.2 The discriminative power of the CBCL.

The discriminative power of the CBCL was tested in the first stage of this study by comparing the scores of nonreferred children with the scores of referred children in two ways.

The first way is the investigation of the cumulative frequencies of the total behavior problem scores and total social competence scores, in order to select the critical score or cutoff point at which the best prediction can be made about the child's referral status.

The second method to discriminate between referred and nonreferred

children is by discriminant function analysis as described in the paragraph on the methods (III.8).

IV.3.2.1 Cumulative frequencies of total behavior problem scores.

As was shown in paragraph IV.3, referred children scored significantly higher than nonreferred children on the majority of the behavior problem items. Also it was found that for the total behavior problem score, clinical status accounted for the largest percentage of variance, showing a stronger relation of total behavior problem score with clinical status than any single item did. Therefore, the total behavior problem score was chosen for discrimination between both groups of children.

Instead of presenting the data either for the whole population or for every age and gender group separately, the populations were grouped for both sexes in age ranges 4-5, 6-11 and 12-16 years. This division by age and gender was employed by Achenbach e.a. (1981), making the comparison of the results of their study with ours possible.

Appendix G shows for each gender and age group the sensitivity and specificity. For the nonreferred population, the cumulative frequencies of the total behavior problem scores are equal to the specificity (percentage of non-cases correctly identified as non-cases). The sensitivity (percentage of cases correctly identified as cases) has been obtained by subtracting the cumulative frequencies from 100. As can be seen from Appendix G, the discriminative power of the CBCL is smallest for the 4-5 year old girls and strongest for 12-16 year old boys.

Table IV.7 shows the misclassification rates at cutoff scores for which the sensitivity and specificity are nearest to equal. The misclassification rate is the number of cases incorrectly identified, expressed as a percentage of the total number of children on whom the CBCL was obtained. The overall misclassification rate for the cutoff scores as reported in table IV.7 is 21.6%.

For comparability with Achenbach e.a.'s (1981) data, the cutoff points for total behavior problem scores including 90% of nonreferred children are reported in table IV.8, together with the cutoff points found by Achenbach e.a. (1981). Except for the slightly higher scores for the 6-11-year-olds, the results are strikingly similar. The misclassification rates found in our study when applying the cutoff scores corresponding with the 90th percentile for nonreferred children to the combined referred and nonreferred samples are shown in table IV.9.

Table IV.7

Misclassification rates at cutoff points for which sensitivity and specificity are nearest to equal. Total behavior problem score on the CBCL.

	Boys				Girls			
		misclassification rate				misclassification rate		
Age (yrs)	cutoff score	false+ positives %	false± negatives %	overall %	cutoff score	false+ positives %	false± negatives %	overall %
4-5	30	26.1	25.0	25.6	25	30.7	31.3	30.8
6-11	33	22.0	20.8	21.4	29	23.9	23.3	23.7
12-16	32	17.3	18.1	17.6	27	18.0	19.4	18.5
4-16		20.9	20.5	20.7		22.8	22.6	22.7

+ false positives = percentage of nonreferred children classified as belonging to the referred sample

± false negatives = percentage of referred children classified as belonging to the nonreferred sample

Overall misclassification across all ages and both genders: 21.6%

Table IV.8

Cutoff points for total behavior problem scores on the CBCL including 90% of nonreferred children. Comparison of American and Dutch data.

Age Group (yrs)	American		Dutch	
	Boys	Girls	Boys	Girls
4 - 5	42	42	40	39
6 - 11	40	37	44	41
12 - 16	38	37	38	36

Table IV.9

Misclassification rates at cutoff points corresponding with the 90th percentile of the cumulative frequency distribution of total behavior problem scores obtained by nonreferred children

	Boys				Girls			
		Misclassification rate				Misclassification rate		
Age (yrs)	Cutoff score	False positives %	False negatives %	Overall %	Cutoff score	False positives %	False negatives %	Overall %
4-5	40	9.8	39.5	23.1	39	9.6	47.9	18.2
6-11	44	9.9	35.4	23.3	41	9.7	43.7	21.1
12-16	38	8.9	26.4	16.6	36	9.3	29.7	15.6
4-16		9.5	33.1	20.9		9.5	38.9	18.6

Overall misclassification across all ages and both genders: 19.9%

IV.3.2.2 Cumulative frequencies of total social competence scores.

Inspection of the cumulative frequencies of the total social competence scores of the referred and nonreferred samples shows that although there are clearly differences between these two samples in their distribution of total social competence scores, there is nevertheless a considerable overlap. Appendix H shows the sensitivity and specificity of the total social competence scores. The better a child is socially adapted the higher its social competence scores will be. Therefore, sensitivity is equal to the cumulative frequencies of total social competence scores for the referred sample, while the specificity is computed by subtracting the cumulative frequencies of the nonreferred sample from 100.

Note that the number of children in each age/gender group is smaller than the corresponding groups for which the sensitivity and specificity of total behavior problem scores was reported. This difference is due to the slightly larger amount of missing data on social competence items compared with the missing data on behavior problem items which are virtually zero. Another factor responsible for lower numbers of children in the 6-11 year-age-group was the fact that more than half of the six-year-olds had not attended elementary school yet.

Table IV.10 shows the cutoff points for total social competence scores including 90% of the nonreferred children found in our study as well as those found by Achenbach e.a. (1981). Again the results are strikingly

similar. Table IV.11 gives the misclassification rates at cutoff points for which sensitivity and specificity are nearest to equal and table IV.12 gives the results at the cutoff points listed in table IV.10, corresponding with the 10th percentile of nonreferred children.

Table IV. 10

Cutoff points for total social competence scores on the CBCL, including 90% of the referred children.

Comparison of American and Dutch data.

Age groups (yrs)	American		Dutch	
	Boys	Girls	Boys	Girls
4 - 5	9.0	9.5	7.5	10.0
6 - 11	15.5	16.0	15.5	17.0
12 - 16	15.5	16.0	14.5	16.0

Table IV. 11

Misclassification rates at cutoff points for which sensitivity and specificity are nearest to equal. Total social competence scores.

	Boys				Girls			
		Misclassification rate				Misclassification rate		
Age (yrs)	Cutoff score	False positives	False negatives	Overall	Cutoff score	False positives	False negatives	Overall
		%	%	%		%	%	%
4-5	11.5	35	35	35	13.0	29	31	29.4
6-11	19.5	38	35	36.4	21.0	41	37	39.8
12-16	18.5	34	32	33.1	19.5	38	35	37.1
4-16		35.8	34.0	35.0		37.8	35.7	37.2

Overall misclassification across all ages and both genders: 36.0%

Table IV.12

Misclassification rates at cutoff points corresponding with the 10th percentile of the cumulative frequency distribution of total social compentence scores obtained by nonreferred children

	Boys				Girls			
		Misclassification rate				Misclassification rate		
Age (yrs)	Cutoff score	False positives	False negatives	Overall	Cutoff score	False positives	False negatives	Overall
		%	%	%		%	%	%
4-5	7.5	12	48	28.0	10.0	11	48	18.6
6-11	15.5	11	60	36.1	17.0	10	68	28.5
12-16	14.5	10	49	26.8	16.0	11	49	22.7
4-16		10.8	54.6	32.0		10.5	58.3	24.8

Overall misclassification across all ages and both genders: 28.8%

Table IV.13

Combined behavior problem and social competence cutoff scores. Distribution of children according to each child's behavior problem score and social competence score.

Category	Nonreferred N	Nonreferred (%)	Referred N	Referred (%)
Outside normal range on both cutoffs	50	(2.7)	382	(32.2)
Intermediate*	287	(15.8)	562	(47.4)
Within normale range on both cutoffs	1484	(81.5)	242	(20.3)

*Intermediate category used for those cases falling within the normal range according to one cutoff criterion but not to the other.

IV.3.2.3 Combined behavior problem and social competence scores.

The preceding two paragraphs described how the discriminant power using total behavior problem and social competence cutoff scores was assessed separately. The relation between total behavior problem score and total social competence score in the combined referred and nonreferred samples without the 4-5-year-olds, expressed as Pearson correlation coefficient, is -0.43 ($P<0.001$). Although reasonable, the relation between the two scores is far from perfect. However, we wanted to know how much the combination of both scores would contribute to the discriminative power of the CBCL. Therefore we divided the cases into 3 categories: (1) those cases who scored outside the normal range according to both cutoff criteria, (2) those cases who scored within the normal range according to one cutoff criterion but not to the other, and (3) those cases who fell within the normal range according to both cutoff criteria. Table IV.13 shows the distribution of cases corresponding with these three categories.

In table IV.14 the misclassification rates are presented. Similar to the procedure followed by Achenbach e.a. (1981), these rates were computed for three different options: (1) using an intermediate category, (2) combining the intermediate category with the category of children scoring outside the normal range according to both cutoff scores, and (3) combining the intermediate category with the category of children scoring within the normal range according to both cutoff scores. As can be seen in table IV.14, using an intermediate category produced the smallest overall misclassification but left 28.2% of the children categorized to this intermediate category. Categorizing intermediate cases as within the normal range showed the highest misclassification rate.

Table IV.14

Misclassification rates for different categorizations of intermediate cases.

Categorization	False positives %	False negatives %	Overall %
Intermediate category used	2.7	20.4	9.7
Intermediate cases considered within normal range	2.7	67.8	28.4
Intermediate cases considered outside normal range	18.5	20.4	18.9

IV.3.3 Discriminant analysis.

As outlined in paragraph III.8, discriminant function analysis was applied to the total behavior problem scores on the CBCL of the combined referred and nonreferred sample, which was divided in half in order to obtain a derivation and cross-validation sample. For each gender/age group of the derivation sample, behavior problem items were entered in the order of their reduction of Wilks's lambda until they no longer significantly ($P<0.01$) reduced lambda. The behavior problem items that significantly reduced lambda are reported for each gender/age group in table IV.15. The items are grouped in decreasing magnitude of their standardized canonical discriminant function coefficient. This coefficient expresses the relative contribution of the item to the discriminant function. Hence, it is an indication of the discriminative power of an item (Lindeman e.a., 1980).

The discriminant functions derived for each of the 6 gender/age groups were used to classify the children in their corresponding cross-validation samples. For each gender/age group the percentages of children incorrectly classified are given in table IV.16. The percentage of referred children incorrectly classified as belonging to the nonreferred sample is 28.1%, whereas the percentage of nonreferred children incorrectly classified as belonging to the referred sample is 12.4%. The overall misclassification rate is 20.1%.

IV.4 Association between behavior problems as reported by parents and teachers

Pearson correlation coefficients were computed for CBCL and TRF total scores of the 94 corresponding items on the two checklists. The results for each gender and for the 4-5- and 6-11-year age groups are listed in table IV.17. The correlation across both genders and all ages ($N=1155$) is 0.34 ($p<0.001$). The implications of the rather low agreement between parental and teacher reports are not discussed at this stage of the study, but will be elaborated in chapter VIII.

Table IV.15

Behavior problem items of the CBCL reducing Wilks'λ significantly ($P<0.01$) in stepwise discriminant function analysis. Items are reported in decreasing magnitude of their standardized canonical discriminant function coefficient.

Age group	Boys		Girls	
	Item	Standardized weight	Item	Standardized weight
4-5 years	61 Poor school work	0.56	19 Demands attention	0.62
	25 Poor peer relations	0.53	84 Strange behavior	0.59
	26 Lacks guilt	0.51	7 Bragging	-0.59
			56[g] Vomiting	0.54
			54 Overtired	0.44
6-11 years	69 Secretive	0.42	112 Worrying	0.53
	26 Lacks guilt	0.39	93 Talks too much	-0.48
	96 Sexual preoccup.	-0.35	61 Poor school work	0.38
	25 Poor peer relations	0.33	25 Poor peer relations	0.35
	112 Worrying	0.29	46 Nervous movements	0.31
	6 Encopresis	0.27	22 Disobedient at home	0.30
	45 Nervous	0.27	9 Obsessions	0.29
			82 Steals outside home	0.26
			103 Unhappy, sad, depressed	0.11
12-16 years	112 Worrying	0.53	25 Poor peer relations	0.58
	61 Poor school work	0.51	112 Worrying	0.43
	81 Steals at home	0.42	100 Trouble sleeping	0.41
	46 Nervous movements	0.38	56[a] Aches or pains	0.40
	8 Can't concentrate	0.32	43 Lying or cheating	0.39
			14 Cries a lot	-0.36
			62 Clumsy	0.28

Table IV.16

Percentage of children misclassified in crossvalidation samples using discriminant functions.

Age groups (yrs)	Nonreferred sample (false positives) Boys N (%)	Nonreferred sample (false positives) Girls N (%)	Referred sample (false negatives) Boys N (%)	Referred sample (false negatives) Girls N (%)
4 - 5	53 (5.7)	54 (16.7)	59 (33.9)	22 (63.6)
6 - 11	157 (12.7)	166 (8.4)	248 (26.2)	114 (29.8)
12 - 16	121 (17.4)	129 (13.2)	141 (17.0)	82 (36.6)
4 - 16	331 (13.3)	349 (11.5)	448 (24.3)	218 (35.8)

Overall misclassification across all ages and both genders is 20.1%.

Table IV.17

Correlations between CBCL- and TRF total scores of the 94 corresponding items.

	Boys		Girls		Both	
Age group	(N)	r	(N)	r	(N)	r
4 - 5	(128)	.27	(142)	.28	(270)	.28
6 - 11	(375)	.35	(424)	.35	(799)	.36

Note: - all correlations are significant ($P < 0.01$)

CHAPTER V
DISCUSSION OF THE FIRST STAGE

V.1 Comparison with other studies

V.1.1 Comparison with Achenbach's data.

The mean total behavior problem scores and the mean total social competence scores for age groups 4-5, 6-11, 12-16 years of our referred and nonreferred samples are compared with the American data (Achenbach e.a., 1983) (see table V.1 and V.2). The similarities between both studies are striking, especially for the nonreferred groups. None of the t-tests showed significant differences for either of the age/gender groups in both referred and nonreferred samples. The relatively low mean total behavior problem score in our 4-5-year-old sample of referred girls, may be attributed to its small size which made it more liable to bias.

The frequency with which each behavior problem was reported in both studies was also compared for each gender and 2-year age interval (except for age 16 which was reported separately). Prevalence rates in our general population sample excluding the 43 children referred to a mental health setting past year (the nonreferred sample, N=2033) are compared with prevalence rates in the nonreferred sample of Achenbach (N=1300). A total of 118 (items) x 7 (age) x 2 (gender) = 1652 comparisons were made between our data and those from Achenbach e.a. (1981). Scores of 1 and 2 were combined and only differences equal to or greater than 10% were noted. Of the 1652 comparisons made, 310 (19%) showed differences of 10% or more. For 172 comparisons the American frequencies were higher and for 138 comparisons the Dutch prevalence rates were higher.

Another question was on what kind of symptoms the two studies differed. First we looked for the association of items with empirically derived Externalizing and Internalizing syndromes (Achenbach 1966, 1978). Of the 172 items on which American children obtained higher scores, 58 (34%) were items associated with the Externalizing syndrome and 75 (44%) with the Internalizing. Of the 138 items on which Dutch children scored higher, 69 (50%) were associated with the Externalizing syndrome and only 44 (32%) with the Internalizing Syndrome (table V.3).

Table V.3

Distribution of items associated with Externalizing and Internalizing syndromes. Number of behavior problem items on which American and Dutch prevalence rates differed 10% or more.

	American +			Dutch ‡		
Sex	Boys	Girls	Both	Boys	Girls	Both
Externalizing	29	29	58	45	24	69
Internalzing	36	39	75	20	24	44
Neither	23	16	39	10	15	25

+American prevalence rates higher than Dutch

‡Dutch prevalence rates higher than American

This tendency for Dutch children to obtain higher scores on symptoms associated with the Externalizing syndrome and lower scores on those associated with the Internalizing syndrome was significant (chi square = 7.45, df=1, $p<0.01$). Especially Dutch boys scored higher on the Externalizing syndrome (chi square = 4.29, df=1, $p<0.05$). Those items that showed the greatest difference are presented in table V.4.

Table V.4

Items on which American and Dutch studies disagreed 10% or more in at least 5 different age/gender groups. Number of disagreements in brackets.

American rates higher			Dutch rates higher		
1	Acts too young	(5)	8	Can't concentrate	(9)
2	Allergy	(9)	10	Hyperactive	(14)
3	Argues a lot	(6)	19	Demands attention	(10)
7	Bragging	(5)	34	Feels persecuted	(7)
9	Obsessions	(5)	42	Likes to be alone	(6)
12	Lonely	(6)	45	Nervous	(9)
29	Fears	(6)	58	Picking	(5)
31	Fears impulses	(8)	88	Sulks a lot	(9)
33	Feels unloved	(7)	90	Swearing	(5)
38	Is teased	(10)	92	Walks, talks in sleep	(5)
47	Nightmares	(5)	98	Thumbsucking	(9)
63	Prefers older kids	(7)	102	Underactive	(5)
71	Self-conscious	(13)	104	Unusually loud	(8)
74	Showing off	(13)			
86	Stubborn	(5)			
94	Teases a lot	(6)			
109	Whining	(6)			
112	Worrying	(11)			

Table V.1

Mean total behavior problem scores of referred and nonreferred samples. Comparison between Dutch and American data.

		Nonreferred					
		Dutch			American		
Age (yrs)	Sex	N	Mean	SD	N	Mean	SD
4 - 5	Boys	153	23.3	13.7	100	24.1	14.2
	Girls	166	21.4	13.8	100	25.2	17.1
6 - 11	Boys	454	23.2	17.3	300	21.7	15.0
	Girls	486	20.5	15.2	300	19.9	14.2
12 - 16	Boys	370	18.6	15.3	250	17.5	15.6
	Girls	388	17.0	14.9	250	16.6	14.1

		Referred					
		Dutch			American		
Age (yrs)	Sex	N	Mean	SD	N	Mean	SD
4 - 5	Boys	124	51.2	27.1	100	59.8	30.1
	Girls	48	44.5	24.1	100	58.8	29.1
6 - 11	Boys	505	56.8	26.7	300	58.9	24.0
	Girls	245	50.1	24.6	300	58.4	26.2
12 - 16	Boys	288	57.4	27.3	250	53.1	24.7
	Girls	175	53.8	28.0	250	55.8	26.3

None of the student's t-tests for independent samples between Dutch and American mean total scores of each age/gender group was significant.

Table V.2

Mean total social competence scores of referred and nonreferred samples. Comparison between Dutch and American data.

		Nonreferred					
		Dutch			American		
(yrs)	Sex	N*	Mean	SD	N	Mean	SD
4 - 5	Boys	141	12.8	4.1	100	12.9	2.5
	Girls	162	14.4	3.6	100	13.6	2.7
6 - 11	Boys	385	20.6	3.9	300	20.1	3.2
	Girls	424	21.6	3.6	300	20.4	3.1
12 - 16	Boys	348	20.2	4.4	250	20.7	3.4
	Girls	363	20.6	4.1	250	20.8	3.4

		Referred					
		Dutch			American		
(yrs)	Sex	N*	Mean	SD	N	Mean	SD
4 - 5	Boys	113	9.5	5.3	100	9.1	4.4
	Girls	42	10.2	5.3	100	10.2	3.8
6 - 11	Boys	406	17.1	5.4	300	15.0	3.7
	Girls	199	19.0	5.0	300	15.2	4.0
12 - 16	Boys	263	15.9	5.4	250	14.8	4.0
	Girls	162	17.5	5.5	250	15.4	4.2

Note: All student's t-tests for independent samples between Dutch and American mean total scores of each age/gender group were not significant.

*The numbers of subjects reported here are smaller than those reported in table V.1 as the result of missing data and because for a number of 6-year-old children not attending elementary school the school scale was missing.

We arbitrarily chose to report only those items on which one of the studies showed disagreement of 10% or more for at least 5 different gender/age groups.

The scores on individual social competence items are not reported in this study because the absolute data are not as informative as the prevalence data on behavior problem items. We nevertheless compared our social competence scores in our nonreferred sample with those reported by Achenbach e.a. (1981). Although both studies did not differ much in mean total scores, they differed in the distribution of the scores across items. Dutch children scored lower on the majority of items on the activities scale, whereas they scored higher on III A Number of organizations and III B Participation in organizations. This may reflect a tendency in Dutch children to prefer peer group activities. Dutch children also scored higher in the total social scale, while the reverse was true for the activities scale on which American children obtained higher scores.

On the school scale items, special class placement was reported for fewer Dutch children, whereas the number of Dutch children who repeated school grades increased from 10-11 years onwards. For 14-15-year-old boys, the percentage of children having repeated grades reached 43, while 37% of girls aged 16 had repeated grades. The American rates of grade repetition did not exceed 19% in any age/gender group. Other school problems (item VII.4) were commoner in Dutch children. ANCOVAs of social competence items showed that younger children obtained higher scores on the total school scale. These data show that Dutch children had more difficulties with their school functioning, especially from age 14 onwards. This suggests that in the Netherlands the transition from the elementary school to secondary school is a very difficult task. A number of causative factors may be involved. First, it could be that the elementary school does not sufficiently prepare children for the next phase. Second, the secondary school system possibly places too high demands upon the children's capacities, and the third factor that may be involved is that the choice of secondary school type in the individual child's case is often not optimal due to either a wrong advice, or to the parents' inclination, despite the correct advice, to choose a level which is too high for their child. In the present educational situation it is possible that the stronger differentiation of the Dutch secondary school system and its consequences for follow-up education and occupation, may press parents to choose a type of school which offers the best perspectives.

More Dutch children seem to be kept in normal classes as compared with American children. If this results in greater rates of failures, it may exert a negative influence on cognitive growth and on the child's self esteem.

VI.2 Comparison with data from other studies

From the prevalence studies listed in table II.1, those studies were selected for comparison that reported the frequencies of parents' ratings

of specific behavior problems in children in the general population with a predominant Western cultural background. Prevalence data of 8 studies were comparable with ours. Items were selected that were reasonably similar to ours. When necessary, our age and gender groups were combined to make comparison possible (table V.5).

Table V.5
Comparison with other studies. Differences in prevalence rates of 10% or more.

Study	Nr. of comparisons	Dutch rates higher: Boys	Dutch rates higher: Girls	Other study's rates higher: Boys	Other study's rates higher: Girls
Cullen e.a. 1966	16	43 Lying 74 Showing off 75 Shy, timid 98 Thumbsucking	43 Lying 56B Headaches 74 Showing off 75 Shy, timid 98 Thumsucking		
Fogelman e.a. 1976	7	10 Hyperactive 44 Nailbiting	10 Hyperactive 44 Nailbiting		
Kastrup 1976	34	17 Daydreams 27 Jealous 29 Fears 43 Lying 98 Thumbsucking	17 Daydreams 27 Jealous 29 Fears 43 Lying 87 Moody 98 Thumbsucking		
Lapouse 1958	9	98 Thumbsucking	98 Thumbsucking	47 Nightmares	47 Nightmares
Miller e.a. 1974	30	10 Hyperactive 19 Demands attent. 27 Jealous 87 Moody 98 Thumbsucking	10 Hyperactive 19 Demands attent. 27 Jealous 87 Moody 98 Thumbsucking	101 Truancy	101 Truancy
Pringle e.a. 1966	20	8 Can't conc. 95 Tantrums	8 Can't conc.	22 Disobeys at home 37B Fighting 56B Headaches 112 Worrying	22 Disobeys at home 37 Fighting 44B Nailbiting 56B Headaches 112 Worrying
Rutter e.a. 1970	30	8 Can't conc. 10 Hyperactive 37 Fights 95 Tantrums 98 Thumbsucking	8 Can't conc. 10B Hyperactive 56B Headaches 98 Thumbsucking	56F Stomach aches 112 Worrying	56F Stomach aches 112 Worrying
Shepard 1971	176	27 Jealous	27 Jealous		87 Moody

Of 322 comparisons made on 38 items, 67 (21%) showed differences of more than 10%. For 50 of these differences our rates were higher and for 17 the prevalence rates in our study were lower. Of the 50 differences reflecting higher rates for Dutch children, 28 were for items associated with the Externalizing syndrome, and 11 were associated with the Internalizing syndrome. In contrast, of the 17 differences reflecting lower rates for Dutch children, only 5 were on items associated with the Externalizing syndrome and 11 associated with the Internalizing syndrome. This tendency for Dutch children to score higher on Externalizing and lower on Internalizing syndromes was significant (chi square = 7.77, df=1, $p<0.01$). Items for which Dutch prevalence rates are at least 10% higher than rates found in two or more studies are: 8, Can't concentrate; 10, Hyperactive; 27, Jealous; 43, Lying; 56b, Headaches; 87, Moody; 95, Tantrums; 98, Thumbsucking. There was only one item (112, Worrying) for which prevalence rates in two or more other studies were 10%, or more, higher than ours.

V.1.3 Conclusions

Of course we must be careful with the conclusions drawn from differences found between prevalence rates of specific behavior problems in different studies, because semantic effects of item wording can be responsible for differences instead of real differences in actual prevalence. Of the 8 studies other than Achenbach's, with which we compared our data, 5 were British. Of the 17 disagreements reflecting higher rates for children in the other studies, 15 concerned British studies. Achenbach e.a. (1981) also found in their comparison with other studies, that British children showed higher rates on items associated with overcontrolled behavior. Hence, the difference between our study and the mainly British studies can at least partly be attributed to qualities of British children's behavior as reported by their parents.

Four items showing higher rates for Dutch children in two or three British studies as well as in Achenbach's study are: 8, Can't concentrate; 10, Hyperactive; 95, Tantrums (only for boys); and 98, Thumbsucking. Concentration problems were found to be strongly associated with referral status in our study. The fact that concentration problems and the related symptom of hyperactivity are more frequently reported for Dutch children compared with children in a number of other studies and the fact that concentration problems are strongly associated with clinical status, suggests that the actual prevalence of these symptoms may be elevated in Dutch children.

Another item, though not associated with clinical status, which is much more common in Dutch children is Thumbsucking. Because semantic effects and subjectivity in the judgement of this behavior are minimal with respect to this item, this finding may reflect Dutch parents' greater tolerance of this ubiquitously occurring habit.

The main conclusions we can draw from the comparisons of our data with those from other studies is that Dutch children tend to obtain higher scores on items associated with the empirically derived Externalizing syndrome reflecting undercontrolled, acting-out, or "problem behavior". Although semantic effects of item wording may be responsible for differences in ratings, these effects are reduced when items are grouped together according to their joint association with a syndrome. Why then do Dutch parents rate their children so much higher on symptoms related to generally "disturbing" behavior ? One tentative explanation is that Dutch parents are more permissive towards this kind of behavior in their child-rearing practices. However, more research is needed to explore this rather unexpected finding, which stresses the importance of the investigation of cultural differences in children's behavior and behavior problems, and the role parental attitude plays in these differences.

V.2 Gender differences in the general population sample

The number of items on which boys scored higher than girls, was nearly three times as great as the number on which girls scored higher than boys. This finding can only partly be explained by the possibility that parents perceived boys as having more problems. The difference in mean total behavior problem scores between boys (22.1) and girls (19.7), though significant in ANCOVA ($p<0.01$), accounted for less than 1% of the variance. T-tests between mean total behavior problem scores of girls and boys computed for each of the age groups 4-5, 6-11 and 12-16 years separately, did not show significant gender differences. It seems more likely that the explanation of the difference in the number of items with a male or female preponderance is that in our sample the behaviors reported for girls are more evenly distributed across the behavior problem spectrum as depicted by the CBCL. Behavior problems in boys are especially associated with the Externalizing side of the spectrum (see table IV.1). This is in accordance with a number of studies in which males are reported to be more aggressive than females (Feshbach, 1970). However, although it is obvious that boys show more socially disapproved behaviors, many girls also show this kind of behavior (see the figures of Appendix B). The fact that mental health referral rates for boys are much higher than for girls, might therefore be caused by qualitative differences in behavior problems on the one hand and by a greater environmental tolerance of certain aggressive behaviors in girls on the other hand.

The fact that many authors find a large male preponderance in childhood psychopathological conditions can be explained by the fact that most surveys to arrive at childhood gender differences are carried out with referred samples (Eme, 1979). However, this kind of bias does not explain a male/female ratio of psychiatric disorders of 2:1 in a community survey such as the Isle of Wight study (Rutter e.a., 1970) as well as in our own (see second stage). Perhaps clinicians are right in that they judge behavior problems in girls as much less serious, although the number of problems in girls as reported by their parents is only slightly smaller than that in boys. However, of the 15 items for which clinical status showed a large effect in ANCOVAs (table IV.5), there was only one item (8, Can't concentrate) on which boys scored higher, whereas for the other 14 items there was no preponderance for either gender. These findings do not support the hypothesis that behavior problems reported for girls are less serious. It may therefore be the case that clinicians are less likely to detect some psychiatric problems in girls. This is contrary to the situation with adults, since clinicians are more likely to detect psychiatric problems in adult women than in men (Goldberg, 1980).

The finding that boys tended to score higher than girls on the majority of aggressive and other socially disapproved behavior items is consistent with other data (Achenbach e.a., 1981; Maccoby e.a., 1980). Cultural as well as biological factors are held responsible for this difference (Maccoby e.a., 1980).

The higher scores for boys on the items Can't concentrate, Hyperactivity

and Impulsivity, are consistent with the higher proportion of boys clinically diagnosed as hyperactive (Weiss, 1979; Werry, 1968).

Enuresis, encopresis and speech problems are all more common in boys than in girls. Gross e.a. (1983) hold normal maturational delay possibly responsible for the sex difference in enuresis. Encopresis and speech problems are also thought to be related to biological maturation (Rutter e.a., 1970). However, physical, emotional and learning factors (Kaffman, e.a., 1977) also play an important and not yet fully understood role in the causation of enuresis.

Items that were more common in girls were mainly related to the Internalizing syndrome or to neither Internalizing or Externalizing syndrome, especially with respect to symptoms associated with somatic functioning: 56F, Stomach problems; 4, Doesn't eat well and 55, Overweight. Rutter e.a. (1970) also reported a higher frequency of stomach aches in girls.

Fears are reported for slightly more girls than boys. Whether this finding, which was also reported by Achenbach e.a. (1981), reflects actual higher levels of anxiety in girls or a stronger inclination in girls to report their anxieties to their parents, cannot be concluded from our data.

Although the items 5, Acts like opposite sex, and 110, Wishes to be opposite sex, are rarely reported, they concern virtually only girls. For both items no significant effects of clinical status in ANCOVAs were found because they were also rare in the referred sample. As with fears, we do not know how strong the influence of social desirability is on certain behaviors in boys and girls with respect to the prevalence rates found for both items.

The effects of gender on social competence items were very small. None of the gender effects exceeded Cohen's criteria for small effects. All significant effects showed higher rates for girls, which is also reflected in the slightly higher mean total social competence scores for girls in nearly all age groups. The largest gender effects were on the scores obtained on the school scale and number of nonsports activities. The effect of gender in ANCOVAs of the social competence items on the school scale favors girls, although there was no significant gender effect on the behavior problem item 61, Poor school work. The difference in rates between boys and girls is greatest at age 8-9, suggesting that boys have greater difficulties in adapting to elementary school during the first few years. This finding is consistent with the much higher referral rates concerning educational problems reported for boys during early elementary years (Baldwin e.a., 1971). Our data on school functioning do not necessarily concern learning disabilities, although the higher rates of these disorders in males probably contribute to the gender difference we found. Rutter e.a. (1970) found that reading difficulties were very much more common in boys than girls and occurred in 4 percent of the children, whereas general reading backwardness was present in 6.5 percent. Many different symptoms may interfere with satisfactory school functioning as a secondary effect. School is one of the main areas of adaptation and a child who fails in emotional, cognitive or social functioning may also fail in school.

V.3 Age differences in the general population sample

The developmental perspective is of importance for the understanding of disordered behavior. A cross-sectional survey such as ours does not provide data on the individual course of behavioral problems. However, the relation of behavior problems to children's age gives us information about particular developmental periods, assuming that the cultural environment does not change so much that it causes a particular age trend.

The results showed that age effects on individual behavior problems in the ANCOVAs were much larger in number and size than the gender effects. Age showed a small but significant effect on total behavior problem score in ANCOVA. As can be seen in Appendix C, the mean total behavior problem scores show a general decline with age as a result of the decrease in the absolute number of behavior problems as the child matures. As the prevalence of individuals with psychiatric disorder does not diminish with age (Rutter e.a., 1974), the greater number of symptoms in young children in the general population, may be caused by the more diffuse character of the behavioral problems in earlier developmental stages. If we look at the age effect on total behavior problem score in the combined clinical and non-clinical sample, the age effect present in the general population sample disappears. The significant ($P<0.001$) group x age interaction reflects the nonreferred children's declining scores with age being offset by the referred children's more evenly distributed scores across ages. The fact that for the referred children behavior problems did not diminish with age, indicates that for these children the parents report a less age-dependent level of psychopathology. It is hard to tell from our data whether this is caused by features of the referred children themselves, showing stable levels of psychopathology irrespective of their age, or whether it is a function of the type and degree of perceived deviance leading to referral at different ages.

If we now take a closer look at the age effects on the rates of individual items, we find that items for which this effect was significant at a $P<0.01$ level are evely distributed among the Externalizing and Internalizing syndromes. In younger children more attention seeking, wild and dependent behaviors are reported. Hyperactivity shows a small age effect in ANCOVA. However, the small size of the effect and the fact that for about 40% of 16-year-olds hyperactivity was reported, dot not support the hypothesis that hyperactivity may be caused by a maturational delay (Routh e.a., 1974). A number of items declining with age, indicating their possible relation with maturation (Gross, e.a., 1983; Rutter e.a., 1970), and showing a preponderance in boys was mentioned in the preceding paragraph. They are speech problems, encopresis, and bedwetting. Daytime wetting was also more common in younger children. Masturbation showed higher rates in younger children, especially males. Fears were reported more often for younger children, whereas worrying was commoner in older children. This shift from fear reactions related to specific stimuli like the dark or animals, towards the more anticipatory aspect of worrying, may be related to the increase in cognitive capacities as the child develops.

Item 70, Sees things that are not there, was reported only for a minority of young boys and girls. Little is known about perceptual problems in young children. These phenomena may be more related to the child's weak reality testing especially when anxious, than to adolescent or adult psychotic phenomena. Care was taken in the scoring of the CBCL that only those perceptions were scored that had a reality character and were not seen by others. The low prevalence rates for this item indicates that it is not a common phenomenon normally associated with young age.

For item 8, Can't concentate, no significant age effect was found. However, the peak for boys 8-9 years as is shown in figure 8 of Appendix B was also found by Achenbach e.a. (1982), although in our study this finding was less outspoken. Compared with Achenbach our prevalence rates for this item were generally higher for all ages. Achenbach concluded that this finding suggests, that since 8- and 9-year-old boys have the highest rate of referral to clinics, attentional problems arising in the course of development may often be involved. Our findings do not contradict this hypothesis. Especially so when we look at the graph (10) of the related problem of hyperactivity, showing that the decline with age ($P<0.001$) is interrupted at ages 8-9.

Next those items on which older children scored higher than younger children will be considered. The items associated with the Externalizing syndrome on which older children score higher differ from those that are more common at younger ages. Aggressive symptoms in younger children seem to take a diffuse, less organized and more attention seeking form such as arguing, bragging, destroying things, disobedience, screaming and showing off, whereas in older children aggression is less overt and more organized. Aggressive behavior in older children manifests itself through behaviors associated with a syndrome called Delinquent (Achenbach, 1978, 1979) or Socialized-Aggressive (Quay, 1979). These items are: 39, Hangs around with children who get in trouble; 101, Truancy; and 105, Uses alcohol or drugs.

Poor school work (61), an item strongly associated with clinical status in ANCOVAs and discriminant analysis of behavior problem items, showed a small effect for age, older children scoring higher. The same-sized age effect was found by Achenbach e.a. (1981). In our study the greater problems in academic achievement for older children was also reflected in the social competence scores. Academic performance, (No) grade repetition, and, Total school scale, were the only school related items showing significant age effects and on all of the 3 items, older children scoring lower than younger children. On the behavior problem item 61, Poor school work, the prevalence rates for boys follow a near-identical course with Achenbach's data, except for a little peak in our rates at ages 14-15. What is evident in both studies is the increase in prevalence with age for both sexes, and for boys a clear increase in prevalence during the first few years of elementary school. As there are no indications that changes in cognitive development can account for the decline in academic performance (or increase in the item Poor schoolwork), the effect may be caused by the increasing standards children have to meet in school. This is especially so during the first few years of elementary school when the higher levels of

activity, which were tolerated much more in kindergarten now have to be replaced by attentive listening and quiet working.

The general increase with age of children for whom item 42, Likes to be alone, is reported and which is nearly exactly opposite to the decrease with age on item 19, Demands attention, shows the increasing capacity and need of the child to function at a greater distance from the parents and to use withdrawal as a means of coping.

The prevalence rates of the items 35, Feels worthless, and 103, Unhappy, sad or depressed follow a near identical age pattern. After a rise in the prevalence with ages 8-9, the prevalence remains at about the same level. For both genders frequencies do not exceed 11% except for girls at 13 years (16.8%), 15 years (11.3%) and 16 years (13.5%) (these rates for single ages not shown in the graphs of Appendix B, because age groups were combined). Our rates are much lower than those reported by Rutter e.a. (1976) in 14-15-year-olds of whom more than 40% reported misery and around 20% self-depreciation in a personal interview with a psychiatrist. Their rates for girls on these items exceeded those for boys. The authors found that the proportion who looked sad at interview was far less than the proportion who reported feelings of misery. A small minority of those who reported feelings of depression were actually diagnosed as depressed. Possibly, the reported feelings represented inner turmoil, which is part of adolescent development rather than an indication of psychiatric disorder. The authors stated that no ready answers were available about the clinical significance of the adolescents' reported feelings, although they found a rise in clinically diagnosed depression from 10 to 14 years. The fact that on our item 103, Unhappy, sad or depressed, the effect of clinical status was second largest, shows that the parental subjective perception of these feelings in their child is an indication of the seriousness of the problem. However, this does not mean that the reported affect is always equivalent to clinically diagnosed depression. It may reflect a common characteristic of disturbed children.

Somatic complaints without known medical cause such as 51, Dizzy and 56b, Headaches appear more often in older children. Although the overall gender difference for the item Dizzy was not significant, the higher rates for the 12-15 year-old girls is clear. Headaches show a more gradual rise up to ages 12-13, and did not show a gender effect. Of course, headaches reported by parents do not consist of one diagnostic entity. Somatic causes unknown to the parents, migraine, and tension headaches for instance, could all have been included within this category on the CBCL. Our data on headaches, however, are consistent with the reports that headache is a common symptom in childhood, being equally distributed among both sexes (Hughes, 1984) and with the finding that the frequency of migraine among 8000 school children increased from 2.5% in age group 7-9 years to 4.6% in age group 10-12 years and 5.3% in age group 13-15 years (Bille, 1962).

The increase of rates with age on item 102, Underactive, was not clear from Achenbach e.a.'s (1981) data. Lapouse e.a. (1964) report higher rates for "physical inactivity" in their 9-12-year-old group than in their 6-8-year-old group. Our rates on "underactivity" show an increase for girls

at ages 8-9 and for boys at ages 12-13. Furthermore the Dutch rates are somewhat higher than rates found by Achenbach.

Age effects on social competence items in ANCOVAs outnumbered gender effects in number and size too. As discussed earlier, younger children obtained higher scores on two school scale items and on the total school scale. The other two items on which younger children scored higher were: II A, Number of nonsports activities, and V 2, Contacts with friends. The higher scores for younger children on number of nonsports activities was consistent with data from Achenbach e.a. (1981) who found this social competence item to be the only one on which younger children scored higher. This finding may reflect younger children's preference for activities in the home rather than outdoor activities like sports. The slightly higher scores for younger children on the quality of contacts with friends was not found by Achenbach e.a. (1981). The age effects on the sports items showed an increase in scores with a maximum around the 11th year followed by a decrease during secondary school period. Secondary school for most children means a rather abrupt change concerning school- and social environment in addition to the emotional development.

Scores on the items concerning organizations and jobs increase with age. Older children also obtain slightly higher scores on total activities scale and total social scale.

V.4 SES differences in the general population sample

Lower SES parents reported more behavior problems in their children than higher SES parents. This can be concluded from the fact that all significant SES effects in ANCOVAs of behavior problem items indicated higher scores for lower SES children. The literature is consistent in its reporting of the tendency for lower SES children to show higher rates of behavior problems. However, the magnitude of the reported effects differs. Cullen e.a. (1966), Lapouse e.a. (1964), Richman e.a. (1982), Rutter e.a. (1970) and Sheperd e.a. (1971) mention only a slight association between increased prevalence of psychiatric disorder and lower SES. However, Achenbach e.a. (1981) and Davie e.a. (1972) found a stronger association between the father's occupational level and behavior problems. Also, social adjustment was found in these studies to be poorer in lower SES children. In another study, Rutter e.a. (1974) found higher behavioral deviance in children whose fathers had semiskilled or unskilled jobs in Inner London. The exact nature of how SES is associated with behavioral deviance is not easy to understand. Many intercorrelated variables may be of influence. The higher scores for lower SES children may be partly explained by the higher rates of stressful life events (especially financial stresses) and with unfavourable type of housing. Both factors were found to be associated with higher rates of psychopathology in children (Richman e.a., 1981; Rutter e.a., 1974). Robins (1979) found evidence that lower SES parents showed less adequate parenting than higher SES parents. Other SES related factors by which psychopathology rates may be increased are language delay in young

children (Richman e.a., 1982) and poor school achievement in elementary school children (Davie e.a., 1972).

Another issue is what type of behavior disorders are related with SES. In table IV.3 it was shown that a large proportion of items on which lower SES children scored higher was associated with the Externalizing or Undercontrolled syndrome. As the CBCL items are rather equally distributed according to their associations with the Externalizing or Internalizing syndromes, it can be concluded that our data support the findings of others that there is a slight preponderance of lower SES children showing more overt aggression. Achenbach e.a. (1981) found a tendency in the same direction and Davie e.a. (1972) found more withdrawal, dependency, hostility and restlessness in lower SES children. Robins (1979) concluded that lower-class children have more antisocial behavior, but they do not show more anxiety or psychosomatic symptoms. In older children, low SES is associated with higher delinquency rates (Remschmidt e.a., 1977). Again the causal mechanisms are complicated and as yet poorly understood. One possible factor is that lower SES parents show more tolerance or permissiveness towards aggressive behaviors in their children.

The effects of SES on social competence items (Table IV.4) support the findings described above. In all social competence areas (activities, social adjustment, and school achievement), upper SES children obtained more favourable scores on the average. However, in contacts with peers, lower SES children were not scored by their parents as less well adjusted than higher SES children.

In conclusion, our data support the results from a number of other studies that lower social class children, as indicated by the parents' occupational level, show more behavior problems as reported by their parents and that they are less competent in areas of social adjustment and school achievement.

V.5 Differences between referred and nonreferred children

For 114 of the 118 behavior problem items, clinical status showed a significant effect at a $P<0.01$ level in ANCOVAs of the normal and referred samples. Achenbach's finding that the largest main effect of clinical status was found for item 103, Unhappy, sad or depressed, was fully supported by our results in which this item showed the second largest clinical status effect. Only 2 of the 15 items (112, Worrying; 84, Strange behavior) in our study for which clinical status showed large effects in ANCOVAs were not among the items showing large effects in Achenbach's study (1981). This finding is a strong indication of the general association of these items wich psychological disturbance in children. Furthermore as Achenbach stressed, changes in items showing the largest clinical status effect (in our study 25, Poor peer relations and 103, Unhappy, sad or depressed) may be especially good indices of improvement in otherwise diverse children.

The strong association in our study between clinical status and poor

peer relations is also reflected by the large clinical status effect in ANCOVAs of the social competence items V.2, Contacts with friends and V1 B, Behavior with other children. The only social competence item showing a larger clinical status effect was VII.4 (No)Other school problems.

Another indication of an item's association with clinical status was the choice of the most significant predictors in discriminant function analysis. Of the 32 items obtained by the 6 discriminant functions across the gender/age groups (see table IV.15) with a positive standardized canonical discriminant function coefficient that significantly reduced Wilks's lambda, there were 20 items showing large effects associated with referral status in ANCOVAs (see paragraph IV.3.1). Of the remaining 12 items, 8 showed medium effects and 4 showed small effects. It should be noted that the ANCOVAs were performed on the whole sample, overshadowing effects that are related to a specific age group and/or gender. However, the variance associated with clinical status in ANCOVAs is a better indication of the discriminative power of each item than discriminant function, since in the last method a number of relevant items are not selected because they correlate highly with another item which was selected.

Most social competence items also showed significant effects of clinical status in ANCOVAs. Item V1 A Number of jobs and IV B, Job performance were the only ones that were scored higher in the clinical sample than in the nonclinical sample, whereas for the total activities scale the clinical status effect was not significant. These findings may indicate that referred children in general are not less competent in a number of activities in contrast to their failing in other important areas such as academic performance and relationships.

V.6 The discriminative power of the CBCL

The discriminative power of the CBCL was assessed in a number of ways. The methods rank ordered according to their overall misclassification rates are:

1. Combined behavior problem and social competence cutoff scores with the use of an intermediate category (9.7%).
2. Combined behavior problem and social competence cutoff scores with intermediate cases considered outside the normal range (18.9%).
3. Behavior problem cutoff score corresponding with the 90th percentile of nonreferred children (19.9%).
4. Discriminant analysis of behavior problem items, with the use of discriminant functions in cross validation samples (20.1%).
5. Behavior problem cutoff scores at cutoff points for which sensitivity and specificity were nearest to equal (21.6%).
6. Combined behavior problem and social competence cutoff scores with intermediate cases considered within the normal range (28.4%).
7. Social competence cutoff scores corresponding with the 10th percentile of nonreferred children (28.8%).

8. Social competence cutoff scores at cutoff points for which sensitivity and specificity are nearest to equal (36.0%).

The combined behavior problem and social competence cutoff scores with intermediate cases considered outside the normal range, was only slightly better than the use of behavior problem cutoff scores at the 90th percentile alone. These figures are comparable to the ones found by Achenbach e.a. (1981), who reported an overall misclassification rate of 17.9% using behavior problem cutoff scores, and Rutter e.a. (1970) who found an overall misclassification rate of 19.2%. Using differential weighing of behavior problem items on the basis of discriminant functions resulted in a higher misclassification rate than with total behavior problem cutoff scores corresponding with the 90th percentile.

The overall misclassification rates using behavior problem cutoffs is highest for the 4-5-year-old samples. This may be explained by the high rates of behavior problems tapped by the CBCL for younger nonreferred children, whereas younger referred children possibly show a different kind of problem for which professional help is sought than older children. In our own department many young children (especially girls) are assessed for developmental and/or physical problems, showing fewer behavior problems than older children. Inspecting the mean total behavior problem scores in table V.1 shows that in the nonreferred sample the scores decrease with age, whereas, in contrast, in the referred sample they increase with age, especially from the 4-5 to the 6-11-year age group. This tendency for the 4-5-year-olds to show higher misclassification rates using behavior problem cutoffs was not found when using social competence cutoffs. In 4-5-year-old girls total social competence cutoffs even resulted in the lowest misclassification rates. Referral problems in young children are thus related more to their lack of adaptive functions and skills than to behavior problems.

CHAPTER VI
POPULATION AND METHODS OF THE SECOND STAGE

VI.1 Background

In the first, extensive, stage of this study emphasis was placed on the determination of the prevalence of specific behavioral problems and social competence as reported by parents of children in the general population and on the identification of differences related to demographic variables. The discriminative power of the instrument used, the CBCL, was put to the test by comparing the scores obtained from large samples of referred and nonreferred children. In the second, more intensive, stage of this study a closer look will be taken at a sample of children selected from the general population sample of the first stage for the following reasons:

1. To provide prevalence data of overall psychiatric dysfunctioning in a representative sample of children from the general population.
2. To provide data on specific clinical psychiatric categories according to DSM III criteria.
3. To investigate factors possibly associated with psychiatric disorder.
4. To relate data on child behavior problems obtained by the Child Behavior Checklist and by the Teacher Report Form to clinical diagnostic assessment procedures.

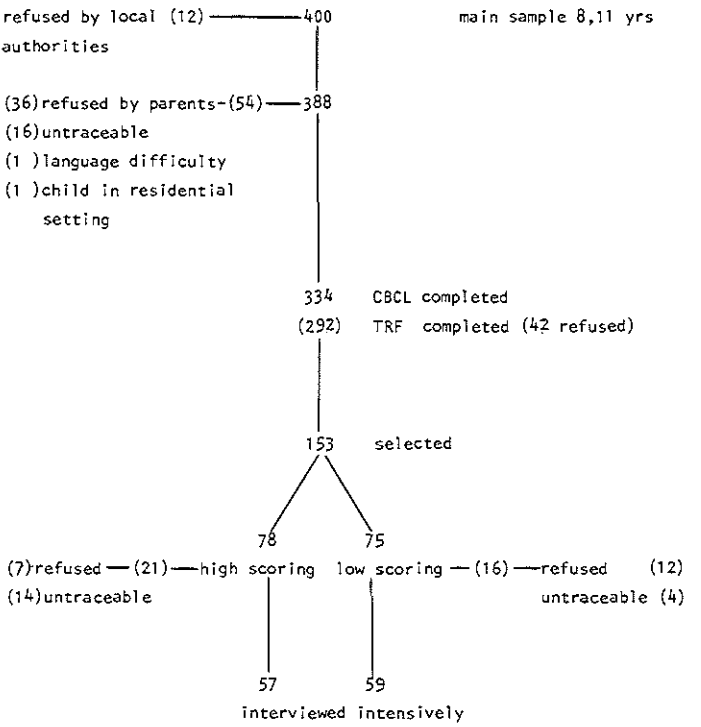
From the literature, especially from the Isle of Wight study (Rutter e.a., 1970), it was predicted that even by using the relatively economic approach of two-stage sampling (paragraph II.3.4), a large number of subjects had to be assessed in order to reach the majority of problem children in the population studied. Due to limitations of available manpower, time and financial means, it was not possible to study the total 4-16-year age range intensively. The choice was made to study the 8- and 11-year-old children in this second stage. The 4-year-old sample was also studied intensively but because of slightly different methodology, the results of this part of the study will be reported elsewhere. The 11-year age group was chosen because it is the age group nearest to the end of the elementary school period for which information was available from the teachers as well as from the parents. The 8-year age group was chosen because at this age it would be clear whether adaptation to elementary

school had succeeded, and because below the age of 8 verbal and attentional abilities were not expected to be adequate for standardized interviews.

VI.2 Selection and composition of the sample

Figure VI.1 shows the selection of the two samples studied. A two-stage sampling procedure was followed. First, as described in paragraph III.1, the CBCL and TRF were obtained for a random sample of the general population of 4-16-year-old children. From the start population of 400 8- and 11-year-old children, no information could be obtained for 12 children owing to municipal authorities refusing to cooperate, 16 children were untraceable, and no information could be obtained for one child due to language difficulties. For one 8-year-old institutionalized child cooperation was refused and in 36 cases parents refused to cooperate. Of the 371 parents reached, 334 (90.0%) completed CBCL's. For 292 (78.7%) children the TRF was filled in by the teacher. The distribution of the response rates for the specific 8- and 11-year age and gender groups are listed in table III.1.

Figure VI.1



In the second stage, children were selected for intensive interviewing, if their total behavior problem score on the CBCL and/or on the TRF was higher than the corresponding scores at the 85th percentile of the cumulative frequency distribution of the total behavior problem score for each of the two age groups. In this way, 78 children scoring in the high- or problem-range on either of the two checklists were selected for further investigation. The 85th percentile was chosen as cutoff score instead of the 90th percentile as described in the first stage, because we wanted to raise the sensitivity of the CBCL and TRF. For the purpose of this study it was considered less serious to interview a number of normal children falsely labeled by the checklist as disturbed, than to miss a number of disturbed children who were not selected by the checklist.

Table VI.1 shows the distribution by age and gender of the children scoring at or above the cutoff points on the CBCL or on the TRF. Except for the 8-year-old children selected by the TRF as belonging to the high scoring group, the children selected by either the CBCL, TRF or both are nearly equally distributed among both sexes. More boys than girls are selected by the TRF as high scorers. However, this difference is not significant (chi square = 1.88, df=1, $p>0.10$). The overlap between children selected as high scorers by the two checklists is only 18.4% for the 8-year-olds and 17.4% for the 11-year-olds.

Table VI.1

Distribution by age and gender of children with total behavior problem scores at or above the cutoff points on the CBCL, TRF or both.

Age (yrs)	8			11		
	n		cutoff	n		cutoff
	Boys	Girls	point	Boys	Girls	point
CBCL	11	7	42	9	7	32
TRF	10	3	45	7	7	33
CBCL and TRF	4	5		4	4	

A normal group was composed by randomly selecting 75 children from the rest of the 8- and 11-year-old children in the general population sample.

The parents of the children selected were sent a letter* at the end of September 1983, explaining the purpose of this second stage of the study. Because one of the purposes of the intensive stage was to obtain information by means of the CBCL and TRF independent from that obtained earlier, the parents were not informed about the group membership of their child (normal group or high scoring group). A separate letter* addressed to the child personally, which also explained the reason for the request, was

*Appendix III, IV

added to the letter meant for the parents.

Parent(s) and child were asked to come to the hospital. Interviews with parents and child were to be carried out separately. The letter addressed to the parents indicated that a traveling-allowance would be provided as well as a small present for the child. A few days after the letter had been sent, the parents were telephoned by one of the research assistents or child psychiatrists, in order to make an appointment in case the parents consented to the interview. Parents for whom no telephone number was indicated, were sent a reply-form and prepaid envelope in addition to the request. On the reply-form they could indicate three alternative appointment dates at their convenience. Parents who did not return the reply-form were sent a reminder in January 1984, followed by a second reminder in February, in case they did not respond to the first one.

Interviews were carried out with 116 children (58 boys and 58 girls) and their parents. Twenty-three parents and children who were willing to cooperate but refused to travel to the hospital were visited at home. Thirty-seven (24.2%) children were not interviewed for a number of reasons: 12 parents could not be reached, neither by telephone nor by mail, after minimally three attempts; 16 parents refused to cooperate without explanation; in 5 cases the parents reported that their child refused to cooperate; 2 parents of mentally retarded children (one child with Down syndrome, and another child unspecified) refused their cooperation because their child was not able to respond verbally; one mother refused to cooperate because her daughter had recently undergone several medical examinations, and one child was in hospital. Parents who did not show up at two subsequent appointments made, were considered to be refusing.

Table VI.2 presents the distribution of the response rate by age and gender. Because it was important to know whether or not the nonresponders were equally distributed among the high scoring and comparison groups, table VI.2 presents the response rates for both groups.

Table VI.2

Response rates (R) for children with total behavior problem scores at or above the cutoff points on either CBCL, TRF, or both and for children scoring below the cutoff points on both CBCL and TRF.

Age (yrs)	8						11					
	Boys			Girls			Boys			Girls		
	n	R	(%)	n	R	(%)	n	R	(%)	n	R	(%)
At/above cutoff	25	17	(68.0)	15	11	(73.3)	20	16	(80.0)	18	13	(72.2)
Below cutoff	15	12	(80.0)	23	19	(82.6)	17	13	(76.5)	20	15	(75.0)
Total	40	29	(72.5)	38	30	(78.9)	37	29	(78.3)	38	28	(73.7)

The differences in response rates between children scoring in the problem range and children from the comparison group are small. The largest difference in proportions was found for the 8-year-old boys. The significance of this difference, tested by applying chi square with Yates' correction for continuity, did not exceed chance probability (chi square=1.01, df=1, P>0.20).

Another approach to investigate whether the nonresponders differed from the responders in level of behavioral deviance is by comparing the mean total behavior problem scores. The mean total behavior problem score on the CBCL for 8- and 11-year-old children scoring in the problem range showed very little difference between the responders (mean 49.6, n=40) and nonresponders (mean 50.6, n=11). On the TRF the responders scored higher (mean 57.9, n=35) than the nonresponders (mean 52.8, n=9). However, a one-tailed T-test showed that this difference was not significant (T=0.92, df=42, P>0.05).

VI.2.1 Discussion

The response rate of 75.8% obtained in the second stage of the study is lower than the 90.0% response rate for the CBCL obtained in the first stage. A number of factors might explain this difference. In the first stage interviewers visited the homes of the children selected, raising the threshold for refusal as compared with the second stage, in which the threshold was lower due to the fact that the parents were approached in a less direct way (telephone and mail). In addition to parents refusing to cooperate, a number of refusals were reported to be due to the child being too frightened to be interviewed. The fact that parents and child were asked to come to the children's hospital has certainly played a role, which can be inferred from the reactions of the children who visited our department. The nonresponders did not differ significantly from the responders in their distribution among high scoring and comparison groups, gender- and age-distribution and in their mean total behavior problem scores. Although the impression was obtained that the nonresponders comprised children who were especially fearful or had medical conditions interfering with normal functioning, it must be realized that this can partly be explained by the sample's composition, because the sample of 153 children contained nearly equal numbers of children scoring in the problem range and children from the comparison group scoring in the normal range. For two children, cooperation was refused because the parents thought that their level of mental functioning made verbal examination impossible.

Although the CBCL and TRF selected the same proportion (15%) of children with total behavior problem scores in the problem range, only 18% of the children were selected by both checklists. This small overlap between children selected on the basis of information from parents and teachers is in accordance with results from other studies. Rutter e.a. (1970) using different questionnaires found an overlap of 16.7%, whereas Mitchell e.a. (1966) found an overlap of about 20%. Whether the difference in children

selected as belonging to the problem group through information from teachers and parents can be attributed to the situation-specificity of the children's behavior, or to variations in the perceptions of teachers and parents, cannot be answered in this stage.

It should be noted that the interviews in the second stage took place about 6 months after the CBCL's had been obtained. This certainly reduces the relationship between the CBCL and the interviews of the second stage because of changes in the child's behavior or of parental attitudes.

VI.3 Methods

VI.3.1 Assessment of the child

For the assessment of the child several procedures were employed. First the Child Assessment Schedule (CAS), developed by Hodges e.a. (1982 a,b) was translated into Dutch. The CAS contains standardized questions, answers, examiner observations and scoring instructions covering the whole range of child psychopathology. The CAS consists of two parts. In the first part the following content areas are covered by 235 specific items: school, friends, activities, family, fears, worries, self image, mood, somatic concerns, expression of anger, and thought disorder symptomatology. The second part consists of 53 items on examiner observations.

Although the authors state that the format of the CAS was designed to enhance rapport with the child, our experience using the CAS in pilot interviews with clinic children was that disturbed children had great difficulties in keeping their attention. Therefore, two breaks were introduced. In the first the child was asked to make two drawings: one by the child's own choice, the other drawing of a person. In the second break the child was asked to make a composition of a standardized play-set of dolls, blocks and other objects. Using the CAS in this way, the instrument was found to reveal a lot of information of the child in a comprehensive way, while retaining the child's attention. Important information not scorable in the CAS was written down in the margin. The total score across items was used by Hodges e.a. (1982 a) as an indication of disturbance. However, instead of relying solely on the child's answers we also scored the clinical impression for each content area (e.g. school, friends, fears, worries, etc.) and for overall functioning, closely similar to Rutter e.a.'s (1970) interview procedure.

Psychiatric functioning of the child was scored by the interviewer directly after the interview had been finished, in the following way: 0, no disorder; 1, ambiguous or trivial; 2, slight disorder; 3, moderate disorder; 4, severe disorder.

Next to the psychiatric interview a number of motor tests from the Yale neuropsycho-educational assessment scales (Shaywitz, 1982) were carried out (results not reported here) and the WISC-R short form (Silverstein, 1972) was administered to obtain an impression of the child's global cognitive functioning.

The interviews were carried out by two child psychiatrists (GB, FV), who were unaware of the scores the child previously obtained on the CBCL and TRF. The children who came to the hospital were interviewed in a room with a one-way mirror, through which a video recording was made. No parents or children objected to this procedure, which was kept as uniform as possible. For the 23 children with whom a home interview was carried out, the interview procedure was kept as closely to the main procedure as possible. Among those children interviewed at home no more were clinically diagnosed as disturbed than among the children who were interviewed in the hospital (chi square=0.83, df=1).

All interviews were carried out with the child alone, while the parents were being interviewed by a research assistant. The child assessment procedure took approximately 75 minutes to 2 hours. All 116 children were able to complete the tasks and to answer the questions. Due to a change of procedure, WISC-R short forms were omitted for 5 children, who appeared to be of normal ability as was supported by their school functioning.

After the interview and the scoring had been completed, the interviewer reported his findings to the parents. We thus took account of the fact that parents and children volunteered in a research program and that the majority did not deliberately seek help for mental health problems. However, when it became clear that the parents perceived a problem in their child and wanted advice, this was given on the basis of our findings. This meant that for some parents explanation, concrete advice or reassurance was sufficient in addition to the proposal that they could contact us again if they wanted to. For instance, the parents of a hyperkinetic 8-year-old boy with attentional problems and until recently an undetected learning disability, were advised to contact their school advisory service. Some parents who were worried about their child's mental health were helped with referral to a mental health agency, which in three cases was our own department, and in others meant that the general practitioner or mental health agency were contacted. In a minority of cases, parents were unaware of a serious problem in their child, which was the case in an 8-year-old depressed boy who reported suicidal thoughts. When these cases occurred, advice was given and accepted.

From both interviewers, 10 videotaped interviews were randomly selected for assessment of inter-interviewer reliability. These interviews were scored by the other interviewer from tape. Intraclass correlation coefficients were computed across the mean scores for each content area and for total score of the CAS (Table VI.3), as well as for the clinical rating of each content area (Table VI.4). For the 0-4 severity scoring of the child's global psychiatric functioning, the ICC between the two interviewers was 0.86, where the kappa (see below) was 0.48 (n=20).

Table VI.3

Intraclass correlation coefficients (ICC) for agreement between two raters for content area scores and total score on the Child Assessment Schedule (n=20)

Content area	ICC	Content area	ICC
School	.96	Self-image	.84
Friends	.97	Mood	.74
Activities	.71	Somatic concerns	.83
Family	.85	Expression of anger	.75
Fears	.87	Observational judgements	.66
Worries	.87	CAS total score	.94

Table VI.4

Intraclass correlation coefficients (ICC) for agreement between two raters for clinical rating of each content area and total score (n=20)

Content area	ICC	Content area	ICC
School	.77	Sleep problems	.73
Friends	1.00	Poor appetite	1.00
Siblings	.74	Pain	.83
Parents	.67	Enuresis	.93
Fears	.89	Encopresis	1.00
Compulsions/Obsessions	1.00	Antisocial behavior school	.59
Worries	.83	Antisocial behavior outside school	.90
Self-image	.80	Observational judgements	.92
Mood	1.00	Total score	.96

VI.3.2 Assessment of the parents

The semi-structured parental interview used by Richman e.a. (1982) for the 8-year-old children was translated into Dutch and slightly adapted for the present study. The interview is a combination of the one developed by Graham and Rutter (1968) and the "Behaviour Screening Questionnaire" (Richman e.a., 1971). In addition to questions on family background, general health, and behavior problems, the interview contains questions on the following areas: psychological and physical symptoms of mother and father; characteristics of parent-child relationship; marital relationship and stress on the family. The parent (usually the mother) is asked for recent (past 6 months), concrete descriptions of behavior, rather than for her attitude or opinions about the child's behavior. The same principle applies to the other areas of the interview. When sufficient information had been obtained, the interviewer scored each item according to operationally defined criteria. The scoring alternatives were 0 for no problem, 1 problem somewhat present, and a 2 when the problem was marked. Although this approach reduced the level of inference, interviewers still had to judge the parental answers.

Both parents completed the translated "Malaise inventory" (Rutter e.a., 1970), a questionnaire derived from the Cornell Medical Index Health Questionnaire, comprising 24 items on anxiety, depression and psychosomatic symptoms. From the Yes-No answers the parents gave to the questionnaire a total "Malaise score" could be computed. Rutter e.a. (1970) reported that the "Malaise inventory" differentiated moderately well between parents with and without psychiatric disorder.

Interviews were completed with 67 (57.8%) mothers, 8 (6.9%) fathers, and 41 (35.3%) both parents. In case the interview was carried out with only one of the parents present, a "malaise inventory" and prestamped envelope were given to the parent with the request to have the other parent fill in the form at home and send it back to us. It should be noted that in 75 (64.7%) cases, the interview was carried out in the absence of the other parent. Information on the absent parent's physical and psychological status is therefore coloured by the other parent's view. The same is true for marital problems and for the relationship between the child and the parent who was not present during the interview.

The marital status of the parents was: 107 (92.2%) married, 1 (0.9%) mother had never been married, 8 (6.9%) divorced. All of the 116 target children lived with one or both biological parents. One child had a twin.

The interviews were carried out by four previously trained research assistants who were unaware of the CBCL and TRF scores. The parent interview took about the same time to administer as the child assessment, namely 75 minutes to 2 hours.

Clinical severity ratings of global content areas were made by three child psychiatrists (GB, JS, FV) independently on the basis of the scoring and relevant information from the interviewers. Psychiatric ratings were made for the following areas: physical and psychological health of mother and father; stress on the family; empathy and criticism shown by mother and

father to their child; quality of marital relationships; psychiatric functioning of the child. Ratings were coded on a 0-4 point severity scale, corresponding with increasing severity. When only two raters agreed upon a score, this score was chosen, and in the few cases in which there was no agreement at all, the raters chose a score in consultation with each other.

Reliabilities were computed for interrater agreement. The intraclass correlation coefficient (see paragraph II.5.2 and III.2.2) across the three raters scoring the child's psychiatric functioning was 0.94 (n=116). The mean kappa (see below) for the three raters was 0.70 (n=116).

In 23 randomly selected cases, interrater reliability was computed between scores by one who interviewed the parents and one who observed. The intraclass correlation coefficient computed for the total scores on the behavior problems section was 0.92.

VI.3.3 Overall psychiatric rating and diagnosis

The parental interview and the information from child assessment were reviewed independently for each of the 116 children by three child psychiatrists (GB,JS, FV). The detailed information which was reviewed contained the assessment scores as well as all relevant information written down by the interviewers. However, the psychiatric ratings previously given on the basis of the child assessment only and given on the basis of the parental interview only were excluded when the overall psychiatric ratings and diagnoses were made. This approach was followed in order to determine the relative contribution of information from respectively parent and child to the final diagnosis. The ratings were put in the same 0-4 severity score format as outlined in paragraph VI.3.1.

For dichotomous analyses, scores of 0,1 and 2 represent children who do not need specialized psychiatric help, and scores of 3 and 4 represent children for whom psychiatric attention is warranted.

In addition to the severity rating, a DSM III (APA, 1980) diagnosis was made. The diagnostic approach we used has been described in paragraph II.2.4.4 as "clinical, diagnostic". The judgement whether a disorder was present was based on the following factors: present functioning of the child in terms of behavior, emotions and relationships; duration and severity of abnormalities in functioning; impact of these abnormalities on the child's development and on the child's functioning in family and community. The cognitive level of functioning was also taken into consideration. Abnormal behavior which corresponded with the child's level of cognitive functioning was not considered a psychiatric disorder.

In order to give an impression of the ratings, two illustrations are given below. For reasons of anonymity the names have been changed.

1.

Julia, 9-years-old at the time of the intensive interview (8-years-old when the CBCL had been obtained), has an older brother. She is in the 4th grade and does well at school according to mother. She has a few close friends and she regularly goes horse riding and goes to ballet classes. She is

physically healthy, although she occasionally wets the bed (once a month). She is described by the mother as a lively, happy girl who has virtually no fears, few worries, and no problems as to her conduct. At the beginning of the direct assessment, she appears a little shy. During the interview she makes the impression of being a sensitive girl who verbalizes well. She describes minor problems concerning her relationship with her parents. Her fine motor functioning is normal. Total IQ=120. Overall psychiatric functioning: 0 (normal); DSM-III diagnosis on axis I: Functional enuresis 307.60; CBCL total score:4; TRF-total score: 37. Both checklists are below the 85th percentile of the cumulative frequency distribution for the 8-year-old age group (normal range). As both mother and Julia herself do not express any distress concerning the bedwetting, the interviewer only briefly mentions the existence of therapeutic measures about which they can get information from the G.P.

2.

David, 11-years-old is the only child from a Dutch father and foreign mother who speaks with an accent. She is an anxious and slightly paranoid woman. The interviewer gets a strong impression that mother's breath has an alcohol odour. David is in 5th grade now and mother reports problems in his concentration at school. Although according to the teacher's judgement, David could have passed to the 6th grade, it was mother who had insisted on his repeating grades. Mother describes David as a very tense and anxious boy, who has virtually no friends. Mother does not allow him to get out because of all the things that could happen to him. He is nervous, bites fingernails, and is worried about a number of events, such as fire and leakage in the house. He has fears concerning thunder and fire- or police-sirens. During the direct assessment he is a very talkative and tense boy, who constantly asks for reassurance. He is concerned about the quality of his drawings and about the correctness of his answers. He makes an unhappy impression. He is afraid of being killed by his friends because they might think that he betrayed them. He feels more or less relaxed in the presence of his father who takes him out to football matches. Father recognizes his only son's distress but until this moment any attempt from his side to seek help was strongly resisted by his wife. David's anxieties, and unhappiness cause a lot of concern and many marital disputes. The parental quarrels make David anxious. David has low self-esteem and he feels guilty after being angry with his mother. He reports that his mother is very nervous and sometimes hits him. He sounds depressed as he reports:"... sometimes I think: Why am I on earth, why do people exist, because we will die anyway?". The fine motor tests are normal; Total IQ=97; overall psychiatric rating: 4 (severe disorder); DSM-III diagnosis on axis I: (1) Overanxious disorder 313.00; (2) dysthymic disorder 300.40; CBCL total score: 34: TRF total score: 47. Both checklists' scores are above the 85th percentile of the cumulative distribution for 11-year-olds.

During evaluation, the distress of the boy was explained to the parents. Unfortunately, although mother was aware of her son's problems, she could not bear the thought that David should need help and talk to someone else about his problems. We offered our readiness to mediate for referral and

left our telephone number in case the parents wanted further help.

Interrater reliability was assessed by comparing ratings and diagnoses made by the three diagnosticians in all 116 cases. As a measure of the interrater reliability for the severity ratings on the 0-4 interval scale intraclass correlation coefficients were computed. The results are presented in table VI.5. If we consider the severity rating scale as a nominal scale, Cohen's kappa's (see below) can be computed. Kappa's are also listed in table VI.5. However, kappa is not directly applicable to the situation in which multiple diagnoses are formulated by multiple raters. This was the case in this study because DSM III allows multiple diagnoses to be made on each of the axes used in this study (axis I and II). Therefore, kappa's were computed for 6 DSM III categories frequently used in this study. The results are given in table VI.6.

Table VI.5

Intraclass correlation coefficients (ICC) and kappa's for agreement between overall psychiatric scores by rater A, B and C. (n=116)

Raters	ICC	Kappa
A and B	.94	.73
A and C	.95	.76
B and C	.94	.75
A, B and C	.93	.75

Table VI.6

Kappa coefficients of agreement for DSM III Axis I diagnosis between raters A, B and C

Diagnosis	Raters			
	A and B	A and C	B and C	A,B and C
1. Attention deficit disorder with hyperactivity	0.86	0.83	0.78	0.83
2. Conduct disorder	1.00	0.50	0.36	0.50
3. Functional enuresis	1.00	1.00	1.00	1.00
4. Oppositional disorder	0.87	0.76	0.74	0.78
5. Overanxious disorder	0.88	0.86	1.00	0.95
6. Separation anxiety disorder	0.71	0.84	0.84	0.74

VI.3.4 Statistical methods used in the analysis

For the assessment of agreement, corrected for chance agreement, between raters for the DSM III diagnostic classes and severity scales, Cohen's kappa for nominal independent scaling was computed, using the following formula: $k = Po - Pc / 1 - Pc$, where Po is the observed agreement and Pc the chance agreement.

Several measures of the child's behavioral deviance were obtained. The relationship between scores derived from the CBCL on the one side, and from the clinical child assessment, the parent interview and the overall psychiatric rating on the other side, was tested by computing correlation coeficients. The same was done for the scores derived from the TRF, as well as for the sum of the CBCL- and TRF scores obtained by converting the scores on both checklists to standard scores (Z-scores) and computing the sum of standard scores. For the relationship between a number of measures and the 0-4 severity scores used for the clinical ratings, Kendall's nonparametric correlation coefficients for ranked data were computed (Nunally, 1967). For the relationship between the scores derived from the TRF and CBCL, Pearson PM correlation coefficients were calculated.

In order to investigate the respective contributions of the parent interview and the child assessment to the final psychiatric rating, multiple regression analyses were performed. The ordinary PM correlation coefficient r designates the degree of linear relation between two variables, and r^2 indicates the proportion of variance in one variable due to the linear prediction by the other. A multiple correlation (R)

designates the degree to which a criterion variable can be predicted by the linear combination of two or more predictor variables, weighed according to their surplus contribution to the linear prediction. The square of multiple R (i.e. R^2) indicates the proportion of variance in the outcome variable that is accounted for by the sum of weighed predictor variables.

VI.3.5 Discussion

The interrater reliability of the clinical ratings made by the two child psychiatrists on the basis of the direct child assessment (ICC=0.86) proved to be good. Rutter e.a. (1970) found a product moment correlation of 0.84 between ratings of two child psychiatrists who independently interviewed the same children on separate occasions. In our study, the second child psychiatrist's scores were based on the video taped interview conducted by the first.

Whereas the ICC of 0.86 found for the interrater reliability described above was satisfactory, the kappa of 0.48 computed for the same scoring was low. As explained in paragraph II.5.2, kappa is a good measure of agreement for nominal scales, whereas ICC is recommended for interval and ordinal scales such as our 0-4 rating scale. Both measures correct for chance agreement. However, kappa treats the scores as independent and mutually exclusive. For instance, when rater A scores a child 0 (normal) and rater B scores the same child 1 (ambiguous or trivial disorder), the extent to which the judges agree expressed by kappa is the same when rater B scores the child 3 (moderate disorder). The ICC on the other hand, is affected by the magnitude of the scores and by their rank ordering. Therefore, ICC is a better reliability measure for the rating scale we used. However, because clinicians are accustomed to kappa as an agreement measure, we reported both kappa's and ICC's for the agreements of our clinical rating interval scales.

The interrater reliabilities between the two child psychiatrists for content area scores on the CAS were on the average satisfactory (Table VI.3). The lowest ICC was obtained on "Observational judgements". This is probably due to the fact that for the interviewer who scored behavior from the video tape, it was more difficult to observe the more subtle behaviors such as facial expression of emotions.

For the clinical ratings of each content area, interrater reliabilities were on the average high. For some content areas such as compulsions and encopresis, this was due to the fact that these rather rare conditions were not difficult to judge as absent in the sample of 20 children when the child simply did not report them during the interview. Low reliabilities were found for the quality of the child's relationship with his parents and for antisocial behavior in school. From the child's answers it was more difficult to judge these areas.

The inter-interviewer reliabilities between two interviewers independently scoring the same parent interview was good (ICC=0.92), indicating that by using the scoring instructions the interviewers agreed

well in the scoring of the answers elicited from the parents.

The interrater reliabilities between 3 raters scoring the overall psychiatric functioning on a 5-point scale, showed good agreement. As described in paragraph VI.3.3 these scores are based on a combination of both direct child assessment and parent interview. Even the kappa's reached values larger than 0.70. However, it should be noted that the three raters all worked in the same department. Having 3 psychiatists from different locations scoring the children would probably have lowered the agreement. The agreement between the three raters of 0.93 (ICC) is of comparable magnitude to the 0.89 (r) found by Rutter e.a. (1970) for the interrater reliability of overall assessment between two raters.

The kappa's found for the DSM III diagnoses indicated good agreement except for the category of conduct disorder. Kappa's were only computed for the most frequently used categories. The category of conduct disorder was the least frequently employed one for which kappa was computed (for one rater only three cases, for the other two raters in two cases).

CHAPTER VII
RESULTS OF THE SECOND STAGE

VII.1 Relationship between measures.

Tables VII.1 and VII.2 show the Kendall's tau correlation coefficients between the problem scores on the checklists and clinical ratings. Of the relationship between the total behavior problem score of the CBCL and the clinical measures, the CBCL was found to correlate strongest with the parental interview. It should be stressed that the parent interview ratings were based on the same source of information as the CBCL, namely the parents. The child assessment can be regarded as a judgement of the child's psychopathology which was fully independent of the scores derived from the CBCL. The overall psychiatric assessment rating, which was based on information from the child interview and from the parent interview, was found to correlate somewhat less with the CBCL scores than the parent interview. Correlations between child assessment and the checklists were higher for boys than for girls, whereas for the parent interview this gender effect was only consistently present in relation to the TRF. As can be seen from the tables, the correlation between clinical measures and the TRF are much lower, and when the samples are broken down according to gender, the relationship between the TRF and clinical measures even show nonsignificant correlations for girls.

The correlations between the summed standard scores of both CBCL and TRF, and clinical measures were on the average somewhat higher than the correlations with both checklists separate. However, for girls the nonsignificant correlations between clinical measures and the TRF scores reduced the correlations with the combined CBCL/TRF.

In the first stage of this study we chose referral status as the morbidity criterion. In this stage we regard the clinical, diagnostic overall rating as the measure of the child's psychopathology. Table VII.3A shows that the correlation of overall psychiatric rating with the child assessment ratings were higher than with the parent interview ratings.

In order to investigate the respective contributions of each of the two intensive investigations (parental interview and child assessment) to the final psychiatric rating, multiple regression analysis was performed with overall psychiatric rating as dependent or outcome variable and the parent

Table VII.1

Kendall correlation coefficients* by age for relations between scores from the checklist and other assessment procedures.

	CBCL			TRF			CBCL + TRF		
Age	8	11	8/11	8	11	8/11	8	11	8/11
n	59	57	116	55	55	110	59	57	116
Child assessment	.37	.46	.42	.27	.32	.28	.43	.48	.46
Parent interview	.53	.49	.50	.30	.28	.28	.51	.50	.50
Overall rating	.37	.47	.42	.34	.28	.29	.45	.45	.46

* All correlations were significant ($P < 0.05$)

Table VII.2

Kendall correlation coefficients* by gender and age for relations between scores from the checklists and other assessment procedures.

	CBCL				TRF				CBCL + TRF			
Age	8		11		8		11		8		11	
Gender**	B	G	B	G	B	G	B	G	B	G	B	G
n	29	30	29	28	27	28	28	27	29	30	29	28
Child assessment	.42	.33	.55	.34	.31	-	.46	-	.46	.44	.66	-
Parent interview	.45	.61	.56	.47	.30	-	.42	-	.43	.61	.66	.37
Overall rating	.33	.40	.58	.32	.31	-	.41	-	.38	.53	.63	-

* Non-significant correlations not indicated ($P > 0.05$)

** Gender: B=Boys, G=Girls

interview and child assessment as the independent or predictor variables. Table VII.3.B shows the results of the stepwise multiple regression analysis. The percentage of variance (r^2) in the overall psychiatric rating accounted for by respectively the parent interview and child assessment are compared with the square of the multiple R, known as multiple regression coefficient. As can be seen, the proportion of variance in the outcome variable is significantly increased by either of the two predictor variables.

The overlap of children selected by the CBCL and by the TRF was only 18% (see Table VI.1). Therefore, it will not be surprising that a Pearson PM correlation coefficient between the CBCL- and TRF total scores of only .24 ($p<0.01$) was found for the 110 children for whom both checklists were filled in. However, not all the items on the TRF and CBCL were exactly the same. Therefore, we excluded the non-corresponding 24 of the 118 items from both instruments and computed the Pearson PM correlation for the remaining 94 items, which was .26 ($p<0.01$; N=110).

Table VII.3.A

Kendall correlation coefficients (W)* by age and gender between overall psychiatric rating and Child- and Parent interview

Age	8				11			
	Child assessment		Parent interview		Child assessment		Parent interview	
	W	(n)	W	(n)	W	(n)	W	(n)
Both genders	.82	(59)	.73	(59)	.87	(57)	.77	(57)
Boys	.84	(29)	.80	(29)	.85	(29)	.72	(29)
Girls	.77	(30)	.62	(30)	.89	(28)	.79	(28)

*All correlations were significant ($P < 0.05$)

Table VII.3.B

Percentage of variance (r^2) in overall psychiatric rating accounted for by parent interview and child assessment, and Multiple regression coefficient (R^2) by gender and age.

Age	8			11		
	Boys	Girls	Both	Boys	Girls	Both
Variance accounted for by:						
Parent interview	71	56	67	65	72	72
Child assessment	77	69	75	81	85	83
Multiple regression coefficient	88	81	84	86	91	89

VII.2 Overall psychiatric rating and the effectiveness of the CBCL and TRF as screening instruments.

The relationship between the total scores on the CBCL and the TRF and the independent overall psychiatric rating (ranging from 0-4) are visualized in figures VII.1-VII.4. The higher correlations of the overall ratings with the CBCL than with the TRF as presented in the preceding paragraph can also be seen from the figures. As outlined in paragraph VI.2 poblem children were selected on the basis of their scores on the TRF or the CBCL. As cutoff score the 85th percentile of the cumulative frequency distribution of the 8- and 11-year-olds in the general population sample was chosen.

The effectiveness of the screening instruments at the chosen cutoff points was expressed in the sensitivity, specificity and overall misclassification rates as measures. These measures are reported in table VII.4 and VII.5 for two different morbidity criteria.

Table VII.4

Sensitivity, Specificity, Misclassification rate (M.R.) for CBCL, TRF and either CBCL or TRF. Overall psychiatric rating of 3 or 4 as morbidity criterion (respectively moderate and severe disorder).

	8 years						11 years					
	CBCL		TRF		Both		CBCL		TRF		Both	
	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)
Sensitivity	59	(22)	45	(20)	77	(22)	63	(19)	47	(17)	79	(19)
Specificity	78	(37)	89	(35)	70	(37)	82	(38)	82	(38)	63	(38)
M.R.	29	(59)	27	(55)	27	(59)	25	(57)	29	(55)	32	(57)

Table VII.5

Sensitivity, specificity and misclassification rate (M.R.) for CBCL, TRF and either CBCL or TRF. Overall psychiatric rating of 4 as morbidity criterion (only severe disorder).

	8 years						11 years					
	CBCL		TRF		Both		CBCL		TRF		Both	
	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)
Sensitivity	70	(10)	50	(8)	90	(10)	100	(5)	75	(4)	100	(5)
Specificity	71	(49)	81	(47)	61	(49)	73	(52)	77	(51)	54	(52)
M.R.	29	(59)	24	(55)	34	(59)	25	(57)	24	(55)	42	(57)

In table VII.4 ratings 3 (moderate disorder) and 4 (severe disorder) are used as morbidity criteria and in table VII.5 only rating 4 is used as the criterion. As can be seen the sensitivity for the CBCL alone is much higher than for the TRF. Using both CBCL and TRF as screening instruments the sensitivity was greatly increased, slightly at the cost, however, of the specificity. The sensitivity went up further when the stricter morbidity criterion of a score of 4 was used. This increased the sensitivity even up to 100% for the 11-year-olds. However, the absolute number of cases is very small when we use the strict criterion. Furthermore, the specificity decreased to 54%.

Table VII.6 shows that 10 out of 11 (91%) children who scored at or above the cutoff points on both checklists, were rated as disordered (overall rating 3 or 4). Disturbed children (rating 3 or 4) scored significantly more above the cutoff points on both checklists than children with an overall rating of 0, 1 or 2 (chi square = 28.26; df=2; $p<0.001$). If children selected on only one checklist are regarded as an intermediate group, the overall misclassification rate would be 14.3%. However, in that case a substantial number of children (40%) in fact would remain unclassified.

Table VII.6

Numbers of 8- and 11-year old children selected by one, neither or both of the CBCL and TRF for the Normal and the Disturbed group.

	Selected on Both	Selected on One	Selected on Neither
Normal (score 0, 1 or 2)	1	24	50
Disturbed (score 3 or 4)	10	22	9

chi square = 28.26, df = 2, $P < 0.001$

If we raise the cutoff scores to the level chosen in the first stage, namely to those corresponding with the 90th percentile, sensitivity is lowered whereas specificity is increased. However, the overall misclassification rates using the 90th percentile differed little from those obtained using the 85th percentile. For the CBCL the misclassification rate for 8-year-old children decreased from 29 to 27% while for 11-year-old children the misclassification rate remained 25% when the cutoff score was raised.

As can be seen from table VII.5 nearly all severely disturbed children were selected by using both checklists. The proportions of children missed by the checklists could be used in order to estimate the prevalence.

Fig. VII.1

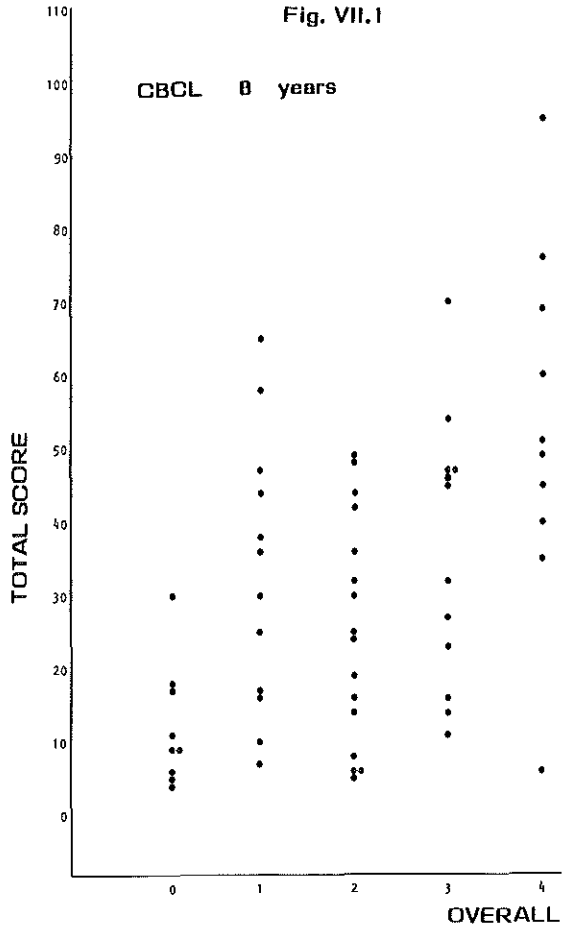


Fig. VII.2

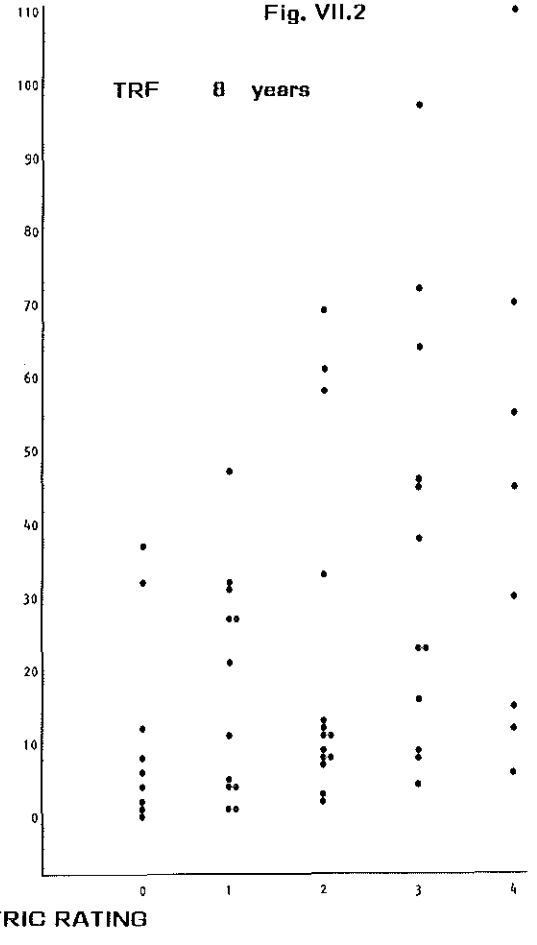


Fig. VII.3

CBCL 11 years

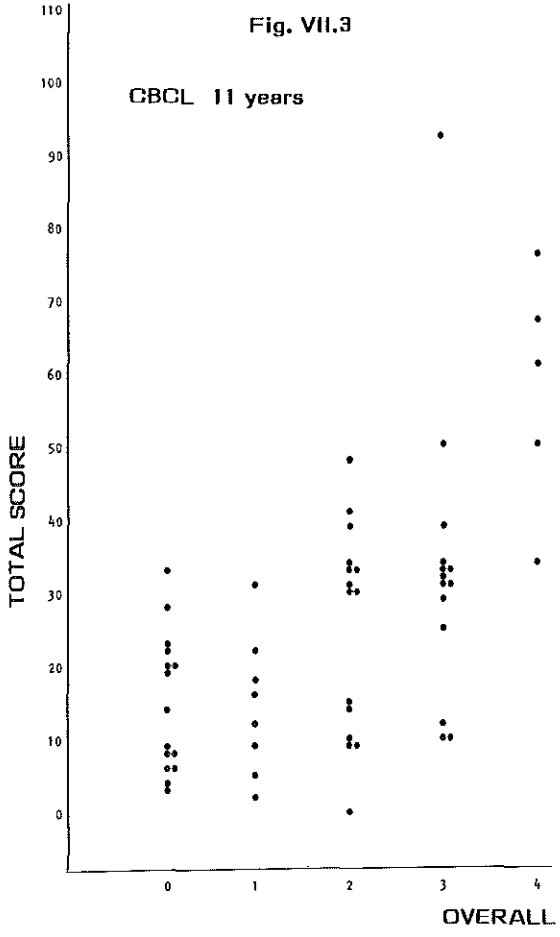
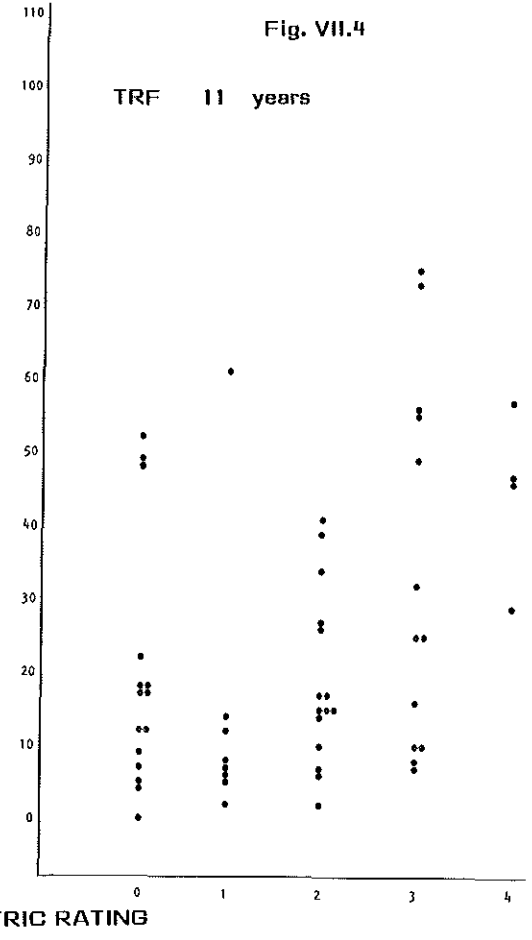


Fig. VII.4

TRF 11 years



The social class distribution between the children designated as scoring high on the CBCL (n=40) and children scoring high on the TRF (n=28) did not show a significant difference. The same was true for the distribution of gender, of children who attended special classes, of children showing academic or other problems in school, and of children who had repeated grades.

VII.3 The prevalence of psychiatric disorder

Table VII.7 shows the results of the overall psychiatric ratings and the number of children scoring above the cutoff scores on either CBCL or TRF (high scoring group) and those scoring below the cutoff scores on both checklists (low scoring group) for the 8-year-olds in the interviewed sample. In order to estimate the prevalence of psychiatric disorder in the total sample, children scoring 3 (moderate disorder) or 4 (severe disorder) were considered to have a psychiatric disorder. It can be seen from table VII.7 that the observed probability of being a case (disordered) given either CBCL or TRF total behavior problem score is above the cutoff point is: 17/28=0.61, whereas the probability of being a case given both CBCL and TRF negative is: 5/31=0.16. In the total sample of 175 8-year-old children there were 40 high scoring children and 135 low scoring children. It can be estimated that in the total sample, the expected number of cases in the high scoring group = 17/28x40=24.3 and the expected number of cases in the low scoring group = 5/31x135=21.8. The estimated prevalence of moderate and severe psychiatric disorders in the total population 8-year-olds for whom CBCL's were filled in is 46/175=26% (95% confidence intervals: 19-33%). However, on 29 children in the total population the TRF was missing, slightly reducing the chance to be selected. The chance that a child with an overall score of 3 or 4 was selected on information of the TRF alone was 4/59=0.07. When we correct for the 2 cases that would have been missed, the corrected prevalence is: 27% (95% confidence intervals: 21-34%). The same procedure can be followed for an overall psychiatric rating of 4 as morbidity criterion. In table VII.8 the distribution of 11-year-olds according to their overall assessment score and the CBCL and TRF scores are given. Table VII.9 gives the estimated prevalence rates for both ages and the two morbidity criteria.

As can be seen, the prevalence rate for moderate or severe disorder is for the 8-year-olds only slightly higher than that for the 11-year-olds. However, the difference in the prevalence rates of severe disorder between the two age groups is much larger. The small absolute number of children scoring 4 on the overall psychiatric rating makes the chance for error in the prevalence estimates greater.

Table VII.7

Distribution of 8-year-olds according to their overall psychiatric rating and their scores on the CBCL and TRF

	Overall psychiatric rating			
	4	3	0,1 or 2	total
CBCL or TRF above cutoff point	9	8	11	28
CBCL and TRF below cutoff point	1	4	26	31

Table VII.8

Distribution of 11 year-olds according to their overall psychiatric rating and their socres on the CBCL and TRF

	Overall psychiatric rating			
	4	3	0,1 or 2	total
CBCL or TRF above cutoff point	5	10	14	29
CBCL and TRF below cutoff point	0	3	24	28

Table VII.9

Estimated prevalence rates (in percentages) of psychiatric disorder in the 8-year-old and 11-year-old total population. (95% confidence intervals in brackets).

		Morbidity criterion	
	N	moderate or severe	severe
8 years	175	27 (21-34)	13 (8-19)
11 years	159	24 (18-32)	4 (2- 9)
8 + 11 years	334	26 (22-32)	7 (4-10)

VII.4 Type of disorder

The disorders of the children obtaining an overall psychiatric rating of 3 or 4 were classified according to DSM III categories. More than one category could fit the child's disorder. Table VII.10 presents the distribution of only the DSM III main-diagnoses, omitting secondary categories. For some combinations of categories it was not feasible to choose a main category. It concerns children showing signs of anxiety or depression together with undercontrolled problem behavior categorized as oppositional disorder. These children were classified as mixed oppositional/neurotic disorder. The number of children in each category was so small that it would make no sense to calculate prevalence estimates of these specific categories for the total population. Mental retardation was classified in one case (a girl with Marfan syndrome) as main diagnosis, because it was not complicated by other psychiatric problems. Two cases, who showed other psychiatric disturbances beside their mental retardation, were classified according to the category of the complicating psychiatric disorder (one case as major depression, and one case, a mentally retarded epileptic girl, as atypical or mixed organic brain syndrome).

Monosymptomatic disorders such as enuresis or speech problems were not rated as psychiatric disorder (received scores lower than 3), unless these symptoms were secondary to more severe psychiatric problems.

Table VII.10

Distribution of psychiatric disorder (overall psychiatric rating 3 or 4) by diagnostic group.

Age	8		11	
Disorder	Boys	Girls	Boys	Girls
Mental retardation		1		
Overanxious disorder	2	1	4	
Separation anxiety disorder		1	2	1
Avoidant disorder			1	
Dysthymic disorder		1	1	1
Major depression	2		1	
Conduct disorder			2	
Oppositional disorder	1			1
Mixed oppositional/other	3		1	1
Attention deficit disorder + H*	7	3	1	
Schizoid disorder			1	
Atypical organic brain syndrome				1
Total	15	7	14	5

*with hyperactivity

The ratio boys:girls was 29/12=2.4. Girls showed no clear preponderance in any category. When we combined the categories, like Rutter e.a. (1970) did, into the broad categories Neurotic, Antisocial, Mixed and Other disorders, we found an excess of boys even for the Neurotic disorders.

VII.5 Comparison between problem and comparison groups

VII.5.1 Behavior in the problem and comparison groups

In the first stage of this study, we used referral status as the morbidity criterion. In this stage we could look at behavior which distinguished problem children from other children from a different angle. We divided the combined 8- and 11-year-old sample into two groups according to their overall psychiatric rating: the problem group consisting of children with an overall score of 3 or 4, and the comparison group consisting of children who scored a 0, 1 or 2.

In order to ascertain which behavior problem items of the parent interview contributed most to the difference between the problem and the comparison group, we looked at the distribution of each item across both groups. Differences were tested by chi-square tests. The rating of the behavior problem items were dichotomized with scores of 1 and 2 combined. The results are listed in table VII.11.

Next, a stepwise discriminant function analysis (for description of method, see paragraph III.2.3) was computed by entering only those items of the parental interview for which a significant ($p<0.05$) chi square between two groups was found. The samples were too small to compose separate derivation and cross-validation samples. The following 5 parental interview items, listed according to their standardized canonical discriminant function coefficient (in brackets), reduced Wilks's lambda significantly:

Unhappy mood	(.50)
Poor relation with sibs	(.40)
Other speech problems (i.e. other than stuttering)	(.40)
Dependency	(.37)
Poor peer relations	(.34)

Using discriminant functions, the percentage of problem children incorrectly classified as belonging to the comparison group is 30.8%, whereas the percentage of comparison group children incorrectly classified as belonging to the problem group is 7.4%. The overall misclassification rate is 15.9%.

Using clinical judgement as morbidity criterion, we wanted to investigate which CBCL items discriminated best between the problem and comparison groups. Therefore, ANOVA was performed on the scores of each item of the CBCL in the problem and comparison groups, in order to assess the effect of psychiatric status. Because the number of subjects (n=116) is small in relation to the 118 CBCL items, we used the magnitude of the effects in ANOVA as a criterion for selecting the number of items for the

discriminant function analysis. Those 27 items for which psychiatric status accounted for at least 6% of the variance (corresponding with a P value of <0.005) and on which discriminant function analysis was performed, are listed in table VII.12.

Stepwise discriminant function analysis selected only 4 items that significantly reduced Wilks's lambda. In the order of their standardized canonical discriminant function coefficient (in brackets) these items are:

9	Obsessions	(.67)
1	Acts too young	(.50)
43	Lying or cheating	(.48)
14	Cries a lot	(.42)

Using discriminant functions, the percentage of problem children incorrectly classified as belonging to the comparison group is 22.5%, whereas the percentage of comparison group children incorrectly classified as belonging to the problem group is 22.7%. The overall misclassification rate is 22.6%.

VII.5.2 Family relationships in the problem and comparison groups

As can be seen from table VII.13, lack of empathy of both parents and overt hostility of mother to the child were much more commonly present in the problem group. Poor marital relationship as indicated by the parent herself or himself was more common in the problem group. The difference increased ($p<0.001$) if instead we took as a measure the judgement of the three psychiatrists about the quality of the marital relationship on the basis of factual information provided by the parent(s) together with observations made by the interviewer of the parent's manner of talking about the marital relationship.

Frequently occurring conflicts between mother and child were also much more common in the problem group than in the comparison group.

VII.5.3 Stress on the family in the problem and comparison groups

During the parent interview, factual information was obtained on the number and the kind of stresses that exerted pressure upon the family in the year prior to the interview and in the previous year. The interviewers rated a number of stresses such as: financial worries, stress concerning work, poor housing, violent quarrels with neighbors, fire, accidents, legal troubles, problems with the other children in the family and health of the family members. On the basis of this information the three psychiatrists, who rated the overall psychiatric functioning of the child, also scored the level of stresses on a 0-4 rating scale. In the same way the physical as well as mental health of the parents were scored. The results of the differences in the problem and comparison groups are shown in table VII.14.

As can be seen, only chronic stresses and mental status problems in the mother were associated with problematic behavior in the children. Of the 15

mothers showing moderate or serious mental status problems, 5 received a diagnosis of depression (2 major depression, 3 dysthymic disorder), and 5 generalized anxiety disorder. None of the 10 specific stresses showed significant differences between the two groups. The proportion of mothers who worked (the majority parttime), was nearly equal in both groups: 42% of the mothers in the problem group and 41% of the mothers in the comparison group.

VII.11

Percentage of children in the problem and comparison groups with specific items of behavior from the parent interview.

Item	Problem group (n=41),%	Comparison group (n=75),%	P‡
Poor appetite	34	24	**
Constipation	15	4	NS
Encopresis	5	-	NS
Enuresis (day)	5	1	NS
Enuresis (night)	20	12	NS
Poor schoolwork	39	15	**
Special class	20	5	*
Extra tutoring	20	5	*
Steals	20	4	*
Fights	27	7	**
Poor peer relations	66	21	***
Concentration problems	27	5	***
Hyperactivity	68	44	*
Nervous	63	33	**
Nervous movements	12	8	NS
Stuttering	12	2	NS
Other speech problems	32	7	***
Difficulty settling at night	34	23	NS
Waking at night	20	12	NS
Dependency	50	16	***
Unhappy mood	71	20	***
Worries	63	35	**
Fears	37	15	**
Difficult to manage	73	36	***
Poor communication	51	13	***
Poor relation with mother	56□	12	***
Poor relation with father	50	13■	***
Poor relation with sibs	63	27	***

‡ Significance level of chi-square test, df=1, *= $P < 0.05$ **=$P < 0.01$ ***=$P < 0.001$ NS=nonsignificant □n=40 ■n=71

Table VII.12

CBCL items selected for discriminant analysis on the basis of the effect of psychiatric status in ANOVA of behavior problems ($P \leq 0.005$; percentage of variance greater than 6%).

Item		v	Item		v
1	Acts too young	7	37	Fighting	15
3	Argues a lot	13	38	Is teased	7
8	Can't concentrate	8	43	Lying or cheating	12
9	Obsessions	21	45	Nervous	11
10	Hyperactive	12	54	Overtired	7
13	Confused	13	62	Clumsy	13
14	Cries a lot	7	65	Refuses to talk	7
16	Cruel to others	11	68	Screams a lot	7
19	Demands attention	7	79	Speech problem	7
20	Destroys own things	11	82	Steals outside home	9
22	Disobedient at home	11	87	Moody	9
25	Poor peer relations	8	95	Temper tantrums	9
35	Feels worthless	15	103	Unhappy, sad	10
			106	Vandalism	7

v = percentage of variance accounted for by effect of psychiatric status

Table VII.13

Percentage of children in the problem and comparison groups with adverse family relationships

Adverse family relationship	Problemgroup (n)	%	Comparisongroup (n)	%	P‡
Conflicts with mother at least daily	(40)	43	(75)	17	**
Conflicts with father at least daily	(40)	13	(70)	4	NS
Mother smacks child > once a week	(40)	3	(75)	-	NS
Father smacks child > once a week	(39)	3	(70)	4	NS
Mother loses control of self with child	(40)	13	(75)	8	NS
Father loses control of self with child	(39)	10	(70)	4	NS
Mother shows little or no warmth to child	(40)	43	(75)	4	***
Father shows little or no warmth to child	(39)	31	(70)	6	***
Mother shows hostility to child	(40)	28	(75)	1	***
Father shows hostility to child	(39)	15	(70)	4	NS
Poor marital relationship	(38)	45	(69)	18	**

‡ Significance level of chi square test, df=1. NS=Nonsignificant

** $P < 0.01$ *** $P < 0.001$

Table VII.14

Differences in problem and comparison groups in stress on the family

Stress	Problem group (n), %	Comparison group (n), %	P+
Acute stress past year	(41) 15	(75) 15	NS
Chronic stress past year	(41) 61	(75) 17	***
Acute stress previous year	(41) 7	(75) -	NS
Chronic stress previous year	(41) 61	(75) 16	***
Health problems mother past year	(40) 5	(75) 3	NS
Health problems mother previous year	(40) 5	(75) 3	NS
Health problems father past year	(39) 15	(70) 7	NS
Health problems father previous year	(39) 10	(70) 9	NS
Mental status problems mother past year	(40) 38	(75) 9	***
Mental status problems mother previous year	(40) 35	(75) 8	***
Mental status problems father past year	(39) 18	(70) 10	NS
Mental status problems father previous year	(39) 18	(70) 9	NS
Mother scores high on Malaise Inventory‡	(40) 30	(75) 7	***
Father scores high on Malaise Inventory‡	(49) 13	(70) 20	NS

+ Significance level of chi square test, df=1 *** $P < 0.001$

‡ Scoring above the 85th percentile of cumulative frequency distribution of the total scores obtained on the Malaise Inventory

CHAPTER VIII
DISCUSSION OF THE SECOND STAGE

VIII.1 Relationship between measures.

Much more than with adults, the child's behavior is characterized by variability, situation specificity and susceptibility to environmental influences. Furthermore, unlike adults, children are rarely motivated to report their problems spontaneously and the cognitive and emotional level of the child's functioning makes verbal reporting more difficult than with adults. Parents are informed about their child across many situations during most of the child's life. Therefore, it seemed justified to treat parents as key informants, as we did in the first stage of this study. However, parental observations may be coloured by subjective factors such as their own tolerance, hostility, denial, or psychiatric conditions. Teachers, on the other hand, can compare a child's behavior with that of other children, especially with respect to social interaction and cognitive functioning. However, the emotional interaction between teacher and child is limited and the child's behavior may be influenced by the classroom situation.

Clinicians might be less biased by emotional factors through their training and professional knowledge. However, it has been found that clinical judgements of whether a child requires psychiatric treatment depend heavily on parental reports and much less on the direct observation of the child (McCoy, 1976). Parental judgement proved to be a good discriminator between referred and nonreferred children in the first stage of this study.

It is well known that the clinician working in an outpatient mental health setting cannot rely solely on short contacts with the child in order to make a diagnosis. The clinican needs information from parents and teachers to form a picture of the child's functioning and developmental perspectives. His efforts are directed towards the integration of all information coming in from different angles in a decision-making process. In this part of the study we investigated relations between information from different sources: the clinician, the parents and the teacher. From a clinical point of view the separation of these different measures of behavioral deviance is rather artificial. Therefore, we also investigated the joint contribution of measures.

The only clinical measure of psychopathology in this study, which was fully independent of the CBCL, was the direct clinical assessment of the child. The interviewers were unaware of the scores that the child had obtained on the CBCL, and the only contact with the parents took place during the few minutes prior to the interview in which the procedure was explained and in which no parental information on the child was elicited.

The level of agreement between the child assessment ratings and the CBCL's total score was influenced by age (higher correlations for the 11-year age group) and by gender (much higher correlation for boys than for girls) as can be seen in tables VII.3.A and VII.3.B. This gender effect is related to the discussion in paragraph V.2, where we saw that in contrast to the clinicians' judging far fewer girls than boys as disturbed, the parental reports did not indicate gender differences in level of psychopathology to such a large extent.

The finding that the clinical assessment ratings for boys showed better correlations with the CBCL total scores than for girls has important implications. For instance, Hodges e.a. (1982 a) reported a correlation coefficient of 0.53 between the total score on the Child Assessment Schedule (CAS) and the CBCL total score. However, their sample of children who were interviewed, consisted of 60 boys and 27 girls, with ages ranging from 7 to 14 years. The authors did not report on the age distribution in their sample. Our findings indicated that studies investigating relationships between parental reports and clinical assessment procedures should test or controle for gender and age effects. If the sample in the Hodges study had contained equal portions of boys and girls, the strength of the relation reported may have been much smaller.

The Pearson correlation coefficient of 0.53 between CAS and CBCL scores as reported by Hodges e.a. (1982 a) comes close to our Kendall's tau coefficient between clinical assessment and CBCL for the 11-year-old boys, which was 0.55, whereas the correlation for girls and for the 8-year-old children in our study is lower. Langner e.a. (1976) reported a correlation of 0.33 between their 35-item total score and the child's direct examination rated on a 5-point-scale for a randomly selected sample of 6-18-year-olds. The correlations between the CBCL's total score and the child assessment ratings found in our study were higher except for the 8-year-old girls for whom we found a correlation of 0.44. Miller (1964) also provided data on the relative value of the following informants: parents, teacher, a psychologist who conducted a family interview, and a child psychiatrist who saw the children in a play session. A correlation of 0.37 between parent and clinician was found by Miller (1964). Our average correlation between CBCL and clinical child assessment was slightly higher (0.42).

The correlations we found between the parental interview and the CBCL were, especially for girls, higher than those found between clinical assessment and the CBCL. This finding is not surprising, since parent interview and CBCL are both based on the parents as source of information.

The parent interview ratings were not equivalent to the direct answers of the parents, as is the case with the CBCL. The interview was scored by

the interviewer. Based on these scorings the clinicians rated the whole interview. In this way clinical judgement played a large role in the rating of the parent interview next to a "retest" effect (the intensive interviews took place about 6 months after the parents had completed the CBCL), which is the reason why the correlation between the interview and the CBCL is not higher.

The overall rating the clinician made on the basis of the information obtained from both the parent interview and the direct child examination showed only for 11-year-old boys higher correlations with the CBCL than either of the two assessment procedures taken separately. The reason why the correlation between overall psychiatric rating and the CBCL is not higher can be explained by the propensity of the three child psychiatrists to rely more on the information obtained from the direct examination of the child than on the information from the parents. This can be inferred from the results of the multiple regression analysis which will be discussed later.

Looking at the agreement between the TRF and the clinical measures it can be seen that the gender differences in correlations are even greater than for the CBCL. For both ages the correlations between TRF and clinical measures were not significant for girls. The correlation between clinical measures and the TRF for boys was only slightly less than the agreement between CBCL and clinical judgement.

In search for an explanation of the low agreement for girls we looked at the clinical diagnoses girls obtained who were selected by the TRF. Of the 17 children with overall psychiatric ratings of 3 or 4, who scored above the cutoff point on the TRF, there were only 5 girls. Two of them were mentally retarded, 2 received a diagnosis of attention deficit disorder with hyperactivity, and the remaining one had a separation anxiety disorder. Of course the absolute numbers are too small to draw firm conclusions. However, it is striking that 4 out of 5 diagnostic categories given to the girls selected by the TRF were associated with behavior interfering with optimal academic functioning. This may suggest that the teachers in our sample were more inclined to select girls with behaviors associated with attentional and/or cognitive difficulties, than those exhibiting behavior problems that do not interfere with academic functioning to a great extent.

The correlation between the TRF and clinical measures was smaller than that between the CBCL and clinical measures. Nevertheless, our correlations (ranging from .27 to .34) were somewhat higher than the correlation coefficient of 0.24 between clinician and teacher as reported by Miller (1964).

The agreement between the CBCL and the TRF was also rather low (0.26). For the 8- and 11-year-olds, this agreement was even lower than for the whole sample of 4-11-year-olds for which a correlation of .34 was found as reported in paragraph IV.4. However, an even lower agreement between parent and teacher checklist was found by Rutter e.a. (1970) who reported a correlation of 0.18.

Combining the TRF and CBCL by summing their respective standard scores

increased the correlation with the overall psychiatric rating and with the direct child assessment except for the 11-year-old girls. For 11-year-old boys a relatively high correlation of 0.66 was found between the combined CBCL/TRF and direct child assessment and a correlation of 0.63 with overall psychiatric rating.

We may conclude from all these findings that the clinician basing his independent judgement only on the direct examination of the child agreed more with parents than with teachers, whereas the agreement between parents and teachers was of the same magnitude as the agreement between teachers and clinicians.

Can we conclude from our results which of the three independent judgements of child psychopathology (i.e. CBCL, TRF, or direct examination of the child) is best ? In our opinion the answer is:"No, we cannot really". Although we might say that teachers showed lower agreement with parents as well as with clinicians, than clinicians did with parents, discarding the teacher information would mean an undesirable loss of important information on a number of children as is shown by the increased sensitivity, when we add the TRF to the CBCL as screening instruments (tables VII.4 and VII.5). It is known from the literature that the agreement between raters who saw children in distinctly different situations is low (Achenbach e.a., 1978). The weight given to reports from different sources varies with the kind of problem the child exhibits and with the aim of the assessment. For instance, biomedical problems might be judged best by the clinician, problems in academic functioning by the teacher, and behavior problems and problems in social functioning by the parents. The following statement made by Achenbach e.a. (1978, p.1289) still holds true and is confirmed by our findings:"...we need far more systematic comparisons of the value of different types of data as guides to prognosis and the prescription of treatment". Our point of view in this stage of the study was that as long as there are no systematically investigated rules for differentially weighing the information from various sources, we used the clinician as the final "integrator" and judge. However, caution should be exercised, if we blindly take clinical judgement as the infallible criterion, since this would prematurely close further investigation.

VIII.2 The contribution of information from parent and child to the overall psychiatric rating

The overall psychiatric rating was based on the parent interview and on the direct child examination. The relation between overall rating and direct child examination was stronger than between overall rating and parent interview. In other words, the relative contribution of the two assessment procedures to the final rating was highest for the direct child examination. However, it was found that both assessment procedures together contributed more to the overall rating than each of them did separately. In the stepwise multiple regression analysis with overall rating as the

outcome variable, both predictor variables remained in the regression equation, which means that the removal of either of them would significantly reduce the percentage of variance accounted for. We may conclude then that both sources of information (i.e. the parent interview and the child assessment) are important contributors to the final diagnosis, with the direct child assessment contributing most. This finding is not in agreement with Rutter e.a. (1970), who found that the personal interview with the parent was by far the most valuable of the three intensive investigations on which the final diagnosis was based (i.e. parental interview, information from teachers, and direct child assessment). They state that the value of the interview with the child was rather in confirming, or refuting, information from other sources, and in providing important information about decisions on the nature and the severity of the disorder.

There are a number of possible explanations for the discrepancy between our findings and those of Rutter e.a. In the first place the 75-minute to 2-hour direct child assessment in our study was much more intensive than the half-hour interview employed in the Isle of Wight study. In the second place the two psychiatrists (GB, FV) who conducted the 116 interviews were among the three psychiatrists (GB, JS, FV) who rated the overall pychiatric status of the child in a later stage, when the detailed information from the parents and from the child could be put together. In the Isle of Wight study the number of psychiatrists who carried out the child interview and the overall rating was larger, although from the publication of 1970 it is not clear exactly how many were involved in the actual interviewing and overall rating. The two interviewing psychiatrists in our study possibly relied more on their own observational judgements based on the intensive interviewing when they made the overall psychiatric rating.

Contrary to what Rutter found, the main impact of the information from the parent interview in our study was in confirming or refuting the impression obtained by the child examination. The interviewers' different theoretical backgrounds may also have played a role in the difference found between the two studies.

VIII.3 The effectiveness of the CBCL and TRF as screening instruments

In the first stage of the study the effectiveness of the CBCL for discriminating between children with pychiatric disorder and those without, was tested by comparing referred and nonreferred children's scores on the instruments in a number of ways. In this stage we did not use referral status as the morbidity criterion but the overall clinical psychiatric rating based on the parent interview and direct child examination. The overall misclassification rates of about 20% found in the first stage when we used total behavior problem cutoff scores on the CBCL and referral status as criterion, were lower than the misclassification rates of 27% for the 8-year-olds and 25% for the 11-year-olds when we used total behavior problem cutoff-scores on the CBCL corresponding with the 90th percentile

and clinical psychiatric diagnosis of moderate or severe disorder as morbidity criterion. Differential weighing of the CBCL items on the basis of discriminant functions yielded a lower misclassification rate of 22.6%. This misclassification rate is higher than the 20.1% found when using discriminant analysis of behavior problem items and referral status as morbidity criterion (see paragraph V.6). The higher accuracy of the instrument we found when tested against referral status as criterion can be explained by the fact that both actual referral and total behavior problem score depended on the judgement and observations of the parents.

The relation between actual referral and direct child assessment is significant but weak (Kendall's tau 0.23, $P<0.05$). Although 7 of the 9 children who actually received professional help within the past 12 months were judged moderately or severely disordered, there were 29 moderately or severely disordered children who dit not receive help.

The question of whether actual referral or clinical judgement is the better morbidity criterion can be answered adequately only if we could have tested both criteria against a number of variables not available in this study, such as long term social adaptation, emotional development, academic achievement, and the ability to stay out of mental health facilities in the future. In the absence of definite answers we can only look for factors that may favour or disfavour both criteria. Referral status owes its value to the fact that it reflects persisting problems of the child in important life areas. Referral is mainly a parental decision not only involving the judgement of the child's problem but also motivational factors in the parents. Parents may report serious problems in their child but not seek professional help. Clinicians base their decisions mainly on the content of the parental report (McCoy, 1976), but are not hampered by factors like denial or motivation which counteract actual referral. However, clinical judgement on need for treatment may well show a tendency in the opposite direction, namely to overestimate this need. The parents in our sample were also asked whether they wanted help for a behavioral or other non-somatic problem of their child. Of the 41 children judged as moderately or severely disordered only 4 parents definitely wanted help but had not recieved it yet for a number of reasons. A large discrepancy remains between the parents' need for help of their child, even if they were not fully motivated or did not know where to get it, and the clinicians' view of whether the child needed it. As far as clinicians are concerned, this discrepancy may be explained by the fact that medical professionals are trained to look carefully for signs referring to illness or psychological problems and they are taught the hazardous effects of missing a serious somatic or psychological condition. In clinical mental health practice, it usually does not harm child and parents when help is offered, even if, strictly speaking, it is not needed. The tendency to overestimate the need for help causes a bias when clinical judgement is chosen as morbidity criterion.

Again our main conclusion is that it is possibly best when effort is put into strategies to elicit the value of different types of data for different questions.

Referral status is a predefined or fixed morbidity criterion. However, with clinical judgement, the morbidity criterion could be varied by including either the severely disordered children only, or both the moderately and severely disordered children. Taking both moderate and severe disorders as morbidity criterion, the sensitivity was low (59% for the 8-year-olds, and 63% for the 11-year-olds), although these percentages were higher than the 54.5% sensitivity found by Rutter e.a. (1970) for their parental questionnaire using the clinical diagnostic approach as morbidity criterion. However, the sensitivity of 53% found for their teacher questionnaire was slightly higher than ours for the TRF (45% for the 8-year-olds and 47% for the 11-year-olds). The specificity for the CBCL and TRF ranged from 78% to 82%. The sensitivity was raised when we used only the stricter criterion of overall psychiatric rating of 4 (severe disorer). This indicated that as the seriousness of the problem of the child increased, there was a better agreement between clinicians on the one hand and parents and teachers on the other. However, this worked at the cost of a lower specificity.

Using both CBCL and TRF as screening instuments raised the sensitivity to respectively 77% and 79% for the 8- and 11-year-olds. The sensitivities come close to the 80% found by Rutter e.a. (1970). Again the sensitivity was raised when the CBCL and TRF were tested against the overall psychiatric rating of 4 as criterion, reaching percentages of 90% for the 8-year-olds and 100% for the 11-year-olds. Of the 11 children scoring above the cutoff points on both the TRF and CBCL, there was only one child who was clinically regarded as normal (false positive), whereas of the 59 children scoring below the cutoff points on both TRF and CBCL there were 9 who were considered disturbed by clinical assessment (false negatives). When both parents and teachers agreed (i.e. either both above or below the cutoff scores) the overall misclassification rate was only 14.3%.

It can be concluded then that the use of both CBCL and TRF as instruments for detecting children with severe psychiatric problems is a sensitive approach at the cost of a substantial number of normal children classified as belonging to the problem group. The use of both instruments is a moderately sensitive approach towards detecting moderate or severe psychiatric problems in children.

When the discriminative power of the CBCL is compared to that of the TRF, the overall misclassification rate of the TRF is slightly lower than that of the CBCL for the 8-year-olds, whereas for the 11-year-olds the overall misclassification rate of the CBCL is slightly better. However, it can be seen that the sensitivity of the CBCL is much higher than that of the TRF.

VIII.4 The prevalence of psychiatric disorder

The comparison of the prevalence rates of psychiatric disorder obtained in this study with those from other community surveys is hazardous, because most surveys differ in their definitions of disorders. The three studies

that are comparable with ours in that they use a two-stage sampling procedure, with a clinical, diagnostic approach in the second stage, are: Leslie (1974) and Rutter e.a. (1970, 1975). The two-stage study of Connell e.a. (1982) was not used for comparison because the authors did not correct for cases missed by the questionnaire. Other studies' prevalence estimates were not comparable because they concerned age groups too far removed from ours.

Leslie (1974) found for her 13-14-year-old children living in Blackburn, U.K. an estimated prevalence rate for moderate or severe disorder of 17.2% and a prevalence rate of severe disorder of 4.4%. Although the author found that her parental questionnaire could have missed cases of moderate disorder, she did not correct the prevalence estimates but concluded that the rates must be considered a minimum. Furthermore, for the screening of the population no teacher information was available, increasing the chance that cases were missed. These methodological factors possibly play a role in the lower prevalence rates Leslie found as compared with our rates of 26% for moderate or severe disorder and 7% for severe disorder in the 8- and 11-year-old population.

Our rate of 26% for moderate or severe disorder comes very close to the 25% found by Rutter e.a. (1975) among 10-year-olds in an inner London borough. These rates are much higher than the 12% reported in the same study for the 10-year-olds in the Isle of Wight. The question was, which factors inherent to the two locations were responsible for the difference in prevalence rates found. The authors explained the area differences by the higher occurrence in the London borough of environmental factors associated with child disorder such as family discord, parental deviance, social advantage and certain school characteristics. Our samples contained 51% children from towns and only 20% living in rural or semi-rural areas (see table III.3). However, the frequencies in our comparison group of marital discord, children not living with both natural parents, psychiatric disorder in the mother, and households with 4 or more children, come close to the respective frequencies of normal Isle of Wight children and are much lower than the respective frequencies for Inner London children. The frequencies of the above factors in our problem group were on the average lower than the ones found for the Isle of Wight as well as for the Inner London population. As we did not have comparable data on the quality of housing and the turnover in school staff and pupils (factors which were less favourable in the inner London area) as well as on socio-economic factors that are possibly of importance such as unemployment, we could not investigate their contribution to the difference in prevalence rates found. Due to a slightly different rating we could not exactly compare the occupational levels of the fathers of children between the other studies and ours.

Although the studies with which we compared our prevalence rates used a clinical, diagnostic approach, it must be realized that the morbidity criteria employed are rather arbitrary and liable to differences related to the investigators' clinical background. This factor might be of crucial importance in considering differences in prevalence rates.

In conclusion, we found that the prevalence rates for our combined 8- and 11-year-old samples are comparable to the ones found in two highly urbanized U.K. areas. However, it seems likely that this finding reflects differences in the rather arbitrary setting of morbidity criteria (i.e. clinical judgement) among groups of investigators.

The effect of age was virtually nonexistent regarding the prevalence of moderate or severe disorder, whereas the prevalence of severe disorder only was much higher in 8-year-olds than in 11-year-olds. Because the proportion of families agreeing to participate with a child scoring at or above the cutoff scores on either checklist was even slightly higher for the 11-year-olds than for the 8-year-olds, differences in response rates between the two age groups could not account for the difference in prevalence found. Of the 15 children receiving an overall psychiatric rating of 4 (severe) there were 10 8-year-olds and only 5 11-year-olds, whereas the sex ratio of 12 boys: 3 girls (4.0) was also very different from the ratio of 29 boys: 12 girls (2.4) found for children receiving an overall psychiatric rating of 3 or 4. Inspection of the DSM-III diagnoses revealed a large preponderance of 8-year-olds receiving a diagnosis of Attention deficit disorder with hyperactivity (6 8-year-olds and only 1 11-year-old). Of the remaining 4 11-year-olds, 2 were diagnosed overanxious and 2 depressed. Of the remaining 4 8-year-olds, 3 were diagnosed depressed and 1 mentally retarded. The age difference found for the prevalence of severe disorders can be attributed to the much higher occurrence of attentional and activity problems in the younger children especially for boys. Hyperactivity was also found to occur significantly more frequently in younger children and in boys in the general population as reported in the first stage of the study (see table IV.3) whereas, for concentration problems the age effect was not linear. Because the combined prevalence of severe and moderate disorders did not differ in the two age groups, it can be concluded that the 3 psychiatrists who rated the children tended to regard attentional problems and hyperactivity as relatively serious problems, despite the fact that the symptom of hyperactivity may be more or less age related. However, only a follow-up of these children can reveal which of them with attentional and activity problems should have been classified as moderately instead of severely disordered.

The samples we studied intensively were too small for deriving prevalence estimates of specific psychiatric diagnostic categories from them. However, an impression of the relative contribution of the specific DSM-III diagnoses to the prevalence of overall psychiatric disorder can be obtained as well as an impression of age and gender effects (table VII.10). Age effect was marked for attention deficit disorder with hyperactivity (more frequent in 8-year-olds, chi square = 8.39, $P<0.001$). The slightly higher occurrence of anxiety disorders in the 11-year-olds was not statistically significant (chi square = 2.82, $P>0.05$).The preponderance of boys among disordered children was marked across all diagnostic groups. The issue was intensively dealt with in paragraph V.2.

VIII.5 Factors associated with psychiatric disorder

VIII.5.1 The child

For the specific items of the parent interview, the association with psychiatric disorder was examined by comparing the frequencies with which these items occurred in the problem group (overall rating of 3 or 4) and in the comparison group (rating of 0,1 or 2). Compared with the first stage of the study, the approach employed here is different with respect to the morbidity criterion (i.e. referral vs. clinical diagnosis) and with respect to the way information was obtained (i.e. self administered CBCL vs. semi-structured parent interview and rating by interviewers). Specific symptoms found to be strongly associated with psychiatric disorder in both approaches can be regarded as important indicators of psychopathology.

The variance associated with psychiatric status in ANOVA is a better indication of the association of each item with psychopathology than discriminant function, because in the latter method a large number of symptoms may be excluded on the basis of their correlation with other symptoms.

An indication of the magnitude of the association of specific symptoms with psychopathology is the persistence of this association across different morbidity criteria. Clinical judgement of psychiatric functioning on the one hand and actual referral status on the other are two distinctly different morbidity criteria for which the association with CBCL behavior problem items was assessed. Of the 27 items for which psychiatric status (overall psychiatric rating 3 or 4) accounted for at least 5.9% of the variance in ANOVA of CBCL behavior problems in the second stage of the study, only the following 4 symptoms were not listed among the symptoms for which referral status accounted for at least 5.9% (i.e. medium or large effects according to Cohen's (1977) criteria) of the variance in ANCOVAs of behavior problems in the first stage of the study: 37, Fighting; 79, Speech problem; 82, Steals outside home; 106, Vandalism. None of the 27 symptoms found in this second stage, which discriminated well between problem and non-problem children were among the least discriminating symptoms in the first stage (i.e. referral status effects accounting for less than 1% of the variance).

The following 7 CBCL items for which referral status showed a large effect in the ANCOVAs described in the first stage (table IV.5), were also among the best discriminating CBCL items in this stage: 8, Can't concentrate; 9, Obsessions; 13, Confused; 19, Demands attention; 25, Poor peer relations; 45, Nervous; 103, Unhappy, sad. It should be realized that the analyses carried out in the second stage concerned only 116 children whereas the effect of referral status was computed on a very much larger sample (N=3420). The fact that the 7 items listed above were also reported by Achenbach e.a. (1981) among the 30 items for which clinical status showed a large effect is a strong indication of their association with psychopathology.

The following symptoms of the parental interview which were highly

associated with psychiatric disorder in this stage ($P<0.01$) were found to be highly associated with referral status as reported in the first stage (clinical status accounted for more than 13.8% of the variance in ANCOVAs, see table IV.6): Poor schoolwork; Poor peer relations; Concentration problems; Nervous; Unhappy mood; Worries; Too fearful.

The question can be raised what practical relevance or consequences these findings have. This study has emphasized through different approaches that children showing one or more of the symptoms described here are in need of further attention. Another question raised as a consequence of our findings might be whether it is necessary to administer the CBCL in its full length or whether an abbreviated version of the instrument can be used equally well. Because this issue is relevant to the study in its totality it will be discussed in the last chapter.

VIII.5.2 The family and the wider environment

It was found that poor marital relationships, parental lack of empathy, maternal hostility, and frequent conflicts with the child were highly associated with psychiatric problems in the child. These results are in agreement with those found by other authors with respect to younger children (Richman e.a., 1982), in the same age range as our children (Rutter e.a., 1974) and older children (Rutter e.a., 1976). It should be noted that the majority of information on environmental factors came from the mother and although Quinton e.a. (1976) found high consistency across the accounts of both father and mother on the quality of their relationship, maternal information on father's empathy and hostility to the child could have been distorted by the mother's perception. Although it is clear from other studies as well as from ours that adverse family relationships are highly associated with psychiatric problems in the child, the direction of the causal chain, if any, cannot be determined from our results. Having a difficult child may evoke negative reactions in the parents, whereas the reverse may also be true, namely that parental hostility or lack of empathy may cause disturbance in their child. Another explanation may be that there is a non-causal relationship between these two factors. A third factor such as a genetic predisposition or some adverse environmental factor may be involved, causing both parent and child to react towards each other in an undesirable way. It is important to realize that the roles of adverse family interactions can be very divergent with respect to different problems.

It was found that in the problem group the amount of chronic stress experiences in the previous year as well as in the year before was much higher than in the comparison group. Likewise mental status problems in the mother but not in the father were associated with psychiatric problems in the children. Except for mental status problems in the mother, none of the other specific stresses on their own were sufficient to be significantly associated with psychiatric disorder in the child. Children who experience a multitude of stresses were found to be much more at psychiatric risk than children with just one risk factor (Rutter, 1979). A combination of chronic

stresses provides much more than a summation of the effects of the separate stresses due to the fact that the stresses seem to potentiate each other. Our data support this finding.

CHAPTER IX
EVALUATION AND RECOMMENDATIONS

In this chapter we will look back at the study in order to assess whether the objectives we set ourselves in the beginning have been reached. Furthermore, the main conclusions and their implications for clinical practice and for research will be discussed, as far as this has not been done in the preceding chapters.

This study, like so many others, has generated more questions than it has answered. Therefore, a selection of recommendations for possible future research will be given.

IX.1 The CBCL, an instrument for clinical practice- and research use ?

The CBCL is nowadays widely used in the United States as an instrument for collecting data on a wide variety of child behaviors that are of clinical concern. The solid quantitative background of this measure, its ready applicability, and the availability of the accompanying Child Behavior Profile, a standardized profile for portraying behavioral disorders and competencies, are major factors regarding the use of the CBCL in child mental health practice as well as in research.

What did this study contribute to the utility of the CBCL ? Because we employed the same general methodology in the first stage of our study as Achenbach e.a. (1981) did, the confirmation of existing validity measures by our data may be regarded a strong indication of the usefulness of the CBCL. In this respect our study is one of the few attempts in developmental psychopathology research to carry out a cross-cultural comparison using the same instrument in a large representative sample of children in a broad age range.

The similarities between the American and Dutch total scores of the nonreferred samples are striking, although the distribution of the scores across some items of the CBCL showed differences.

The content validity criterion of the level of association of CBCL items with referral status showed very similar results in our study to Achenbach e.a.'s (1981) study. Nearly all of the 15 items in our study for which clinical status showed large effects (see table IV.5) were also among the items showing large effects in Achenbach e.a.'s study. In our study, for 4 items no significant differences between the nonreferred and referred samples were found, whereas Achenbach e.a. (1981) found nonsignificant differences for only 2 items. For clinical purposes, this implies that items most strongly associated with referral status in both studies, such as Unhappy, sad or depressed, Nervous or Poor peer relations are good parameters of the child's functioning.

Another validity criterion, in which both studies showed comparable results, is the ability of the CBCL to identify children whose behavior causes so much concern as to warrant professional help. Using total behavior problem cutoff scores corresponding with the 90th percentile of the nonreferred samples as the critical score on which the child's referral status can be predicted, revealed total misclassification rates of 17.6% and 19.9% for the American and Dutch samples respectively.

We have elaborated on the problem of relationships between measures from different sources (i.e. the clinician, the parents, and the teacher) in the preceding chapter. Although our correlation of .42 between the CBCL total score and independent clinical child assessment was among the highest reported in the literature between child assessment and parental information, it is nevertheless rather low. However, clinical child assessment cannot be regarded as a validity criterion for testing an instrument like the CBCL. Clinical child assessment simply provides different information on the child. Whether direct assessment is better than a parental report or not (assuming both assessment procedures are equal with respect to other validity criteria), should be tested against other criteria, such as response to therapeutic agents or measures, long term social adaptation, and school achievement.

In conclusion, we hold the view that the results of this study strongly support the validity of the CBCL as an instrument for collecting data on a wide variety of child behaviors that are of clinical concern in a broad age range.

IX.1.1 Implications

The similarities between the American and Dutch total scores and the demonstration of the usefulness of the CBCL in totally different samples of children in another country, may lead to a self-evident, but nevertheless important conclusion, namely that they support a research strategy to distribute future research tracks, residing within certain main issues, among researchers in different locations. For instance, more systematic

research is needed to assess the relative value of different measures of child psychopathology for different, specific conditions. The same applies to research concerning the relative value of different morbidity criteria. Instead of every researcher or research group working separately, research in developmental psychopathology could progress more, if efforts were combined for tackling related issues in a number of places at the same time. Through the demonstration in this study of the usefulness of the CBCL across rather divergent populations, an important condition has been fulfilled to follow this approach.

As warranted by the results of this study, our research group is continuing the development of the Dutch version of the Child Behavior Profile and is at present undertaking a follow-up project of children on whom information has been obtained for this study.

IX.2 The CBCL (and TRF) as screening instrument

A screening instrument must be able to indicate accurately the existence of a condition in a relatively fast and inexpensive way and should be easily convertible into a score. For instance, a school-doctor who wants to obtain an impression on a child's mental functioning besides information about its physical health could benefit from a screening instrument designed to measure the level of psychopathology. We do not advocate the mass screening of whole populations in order to select disordered children for a number of reasons of both practical and ethical nature, which will not be elaborated here. This study, for instance, showed the impracticality of mass screening, because mental health agencies are simply not able to handle the 26% moderately or severely disturbed children in the general population, apart from the fact that a number of parents will not accept help. However, for a number of purposes the availability of a time saving screening instrument for detecting child mental health problems is of crucial importance.

The power of the CBCL to discriminate between children who exhibit mental health problems and those who do not was tested against two morbidity criteria (i.e. referral status and clinical judgement) and by two methods (i.e. total score cutoff levels and discriminant function analysis). With referral status as morbidity criterion and using the combined behavior problem and social competence cutoff scores corresponding respectively with the 90th and 10th percentile of the cumulative frequency distribution of the normal samples, a misclassification rate of 18.9% was found when categorizing intermediate cases as outside the normal range (the percentage of false positives being 18.5 and false negatives 20.4). Weighing the behavior problem items in a discriminant function did not improve the discriminative power of the behavior problem section of the CBCL. Depending on the purpose for which we want to use the CBCL as screening instrument, the cutoff levels can be varied resulting in a change in the sensitivity opposite to the change in the specificity.

The use of clinical judgement as morbidity criterion showed moderately good sensitivity for the detection of severe disorders. Of the 15 cases diagnosed as severely disturbed by clinicians, 3 (20%) were missed by the CBCL using the 85th percentile of the cumulative frequency distribution of the normal sample. However, numerous "normal" children are falsely categorized as disturbed, resulting in an overall misclassification rate of 27%. Because the social competence frequency distributions had not been available yet when we had to select subjects for the second stage of the study, we could not fully utilize the discriminative power of the CBCL.

When clinical diagnosis of moderate or severe disorder is chosen as morbidity criterion, the sensitivity is lowered, whereas the specificity is increased, although the overall misclassification rate of 27% is equal to the one obtained when only severe disorder was taken as criterion. Weighing the behavior problem items in a discriminant function improved the discriminative power of the CBCL (misclassification rate of 23%). However, our sample size was not large enough for cross validation, which may have increased the misclassification rate.

For service oriented health care, referral status is a more suitable morbidity criterion than clinical judgement, because referral status reflects persisting problems on the part of the child, which are of major concern to the parents, even if they do not perceive problems in the first place.

For school health care services, school guidance services, admission and supervisory committees for special education, as well as for public mental health services, the combined behavior problem and social competence total scores of the CBCL are a moderately accurate measure of the resemblance of a child to those in the referred or nonreferred groups. Especially when the instrument is used within a clinical context to provide a standardized, normative framework of the parent's view of their child next to other sources of data, the CBCL, owing to its self-explanatory design, is a valuable and easily administered measure of a child's mental health problems. However, it must be realized that about 20% of the problem children are missed by the CBCL. In service oriented practice and even in more prevention-oriented service such as in school health care, the CBCL scores have to be placed in a somewhat broader context, for instance by own observations or by interview with the parents before any management decisions are made, such as referral.

In case the number of false negatives need to be minimized, as was the case in the second stage of our study where we wanted to catch as many of the problem children as possible, the CBCL can be complemented with the TRF. In this way the sensitivity increases at the cost of the specificity. For the detection of severe disorders the sensitivity increased from 80% (CBCL only) to 93% (CBCL and TRF), whereas the specificity dropped from 66% (CBCL only) to 52% (CBCL and TRF). For moderate or severe disorders, the sensitivity increased from 61% (CBCL only) to 78% (CBCL and TRF), whereas the specificity decreased from 80% (CBCL only) to 67% (CBCL and TRF). For the detection of severe disorders it is possible to use the CBCL only, whereas for the detection of moderate disorders the CBCL should be combined

with the TRF.

IX.2.1 CBCL full length, or short form ?

For some purposes it may be feasible to reduce the scoring time of an instrument or to make complicated instruments easier to scan. Although the CBCL is not complicated and can be rather easily filled in by parents themselves, the advantages and disadvantages of a reduction of the CBCL will be considered.

Looking at the results of the discriminant function analyses described in chapter IV (see table IV.15), we seem to require only those items selected by the procedure to categorize children as being members of the normal or problem group. However, the overall misclassification rate using the weighing of items in a discriminant function and referral status as criterion, was found to be nearly equal to the misclassification rate when using the total behavior problem cutoff score. With clinical judgement as morbidity criterion, the discriminative power of the CBCL using discriminant function analysis was better than with the use of total behavior problem cutoff scores.

For large scale home surveys it might be advantageous to use discriminant function weighings for categorizing children as being within or outside the normal range. However, there are a number of disadvantages in using discriminant functions. In the first place, while total scores are obtained in the same way across all ages and both genders, the number and kind of items selected by discriminant function analyses vary among the age/gender groups. In the second place, reducing the instrument also reduces a lot of important information, which is undesirable if the CBCL is to be used for more than just the categorization of children as being within or outside the normal range. Furthermore, a reduced CBCL cannot be scored on the Child Behavior Profile. This argument is not yet relevant to the Dutch situation, because at present we do not have profiles available yet with norms for the Dutch population. In future we will construct a Dutch version of the Profile.

IX.2.2 Implications

The discriminative accuracy of the CBCL permits us to use it as an indicator of psychopathology. However, as there still is a misclassification rate of 18.9% found in our study and of 15.5% found by Achenbach e.a. (1981), any management decision should be made in connection with information obtained from different sources.

When no other information is needed than the categorization of children as being within or outside the normal range, differential weighing of items on the basis of discriminant functions may be used. However, for other purposes the reduction of the CBCL is regarded as an undesirable loss of information and in our opinion there are very few reasons why we should not

use the instrument in its full version.

If detection of only severe child mental health problems is needed, then using the CBCL is enough. However, if moderate disorders have to be detected accurately too, the CBCL should be complemented with the TRF. When tested against clinical judgement as morbidity criterion, the TRF was a much less accurate indicator of psychopathology than the CBCL.

IX.3 The prevalence of specific behavior problems - Implications

Differences between prevalence rates of specific behavior problems obtained in other studies and ours were investigated. As stressed earlier we must be careful to draw conclusions from differences in parent reported child behavior across different studies, because semantic effects of item wording, especially when translated into another language, may be responsible for differences. However, differences found between American and Dutch data at the level of the association of items with the Externalizing and Internalizing syndromes permit us to draw conclusions with more certainty.

The fact that Dutch parents rate their children higher on symptoms related to generally disturbing behavior (e.g. Externalizing symptoms) than parents in other studies, cannot be explained by the results of this study. If this finding is confirmed by future research, the next problem to be investigated is whether Dutch parents perceive more undercontrolled behavior in their children or whether Dutch children really behave differently ? If the last question is answered positively, it would be interesting to investigate further the nature of these cultural differences in children's behavior. Furthermore, the relation between culturally determined parental attitudes and children's behavior can give us information on environmental influences on the child's behavior.

For the association of our items with the Externalizing and Internalizing syndromes, we used the factor analytic data derived from American samples (Achenbach e.a., 1983). Factor analysis of samples of referred Dutch children, to be carried out soon, will reveal a more accurate picture of the association of our specific items with the Externalizing and Internalizing syndromes.

IX.4 School failure - Implications

Very striking was the finding that the percentage of children in the nonreferred sample for whom grade repetition was reported, strongly increased after the transition from elementary to secondary school, whereas this was not the case for American nonreferred children. Further investigation of the causes and consequences of this rapid increase in schoolfailures is recommended.

The transition from elementary to secondary school seems to be a stressful event, especially for Dutch children, and it seems warranted to

investigate its effects on further emotional and academic development.

Another finding was the smaller number of Dutch children for whom special class placement was reported, as compared with the American population. It may be important to investigate whether or not children with learning or behavior disorders benefit from a policy to keep them in regular classes.

IX.5 The effect of demographic variables on the prevalence of behavior problems

The effects of gender, age and SES on the prevalence of behavior problems in ANCOVAs were numerous. Of these demographic variables, age showed the largest effect.

Most of the age effects for specific behavior problem items, as well as for total behavior problem score, showed a decline with age in the general population sample. However, in the referred sample, no age effect was present, reflecting either stable levels of psychopathology irrespective of age, or a function of the type and degree of perceived deviance leading to referral at different ages. Many of the relevant age effects were discussed in chapter V. Summarized, the main conclusions were:

1. In general, more attention seeking, wild and dependent behavior is reported in younger children.
2. The following items, reported to be related to maturation, were also more common in younger children in our study: speech problems, encopresis and bedwetting.
3. Aggressive symptoms in younger children were found to be less organized and relatively more aimed at seeking attention, as compared to older children in whom aggression took a less overt and more organized form.
4. Academic performance showed a decline with age, probably reflecting the more exacting standards children have to meet in school.
5. The reported prevalence of the symptoms Feels worthless and Unhappy, sad or depressed, both reflecting problems of affect, showed an increase with ages 8-9, and remained at that level in adolescence. The dramatic rise in symptoms related to depression in adolescence as reported by others (e.g. Rutter e.a., 1976) was not confirmed by our parental reports.
6. Somatic complaints without known medical disorder, such as Dizziness and Headaches, showed an increase with age.

Gender effects were less numerous. Although boys were reported to obtain significantly higher mean total behavior problems scores than girls, this gender effect was small and T-tests performed on separate age groups did not reveal significant gender differences. This finding contradicts the preponderance of psychiatric disorders as diagnosed by clinicians in boys found in community surveys such as the Isle of Wight study (Rutter e.a., 1970), including the second stage of our study.

Boys scored higher on socially disapproved behavior and on concentration and hyperactivity problems than girls. Also, behavior problems thought to

be related to maturational delay such as speech problems, encopresis and enuresis, were more common in boys than in girls.

Girls were found to obtain higher scores on problems relating to Internalizing behavior or to neither Internalizing nor Externalizing behavior. On the level of specific items, girls scored higher on symptoms associated with somatic functioning and on fears. For symptoms like fears and opposite sex behavior, which was also reported more often for girls, it is not clear how strong the influence of social tolerance on the reporting of these behaviors is.

SES was the third demographic variable for which the effect on prevalence rates was investigated. The main conclusion was that for lower SES children more problems and fewer social competencies were reported in the first stage of the study. Furthermore, most of the items reported more frequently for lower SES children were externalizing or undercontrolled behaviors. The parents' occupational level was chosen as indicator of SES level in this study. Rutter e.a. (1974) found child psychiatric disorder to be positively associated with a number of social parameters such as: large family size, labouring or semi-skilled manual job of the father, and living in houses rented from the local authority. Robins (1979) found that lower SES parents showed relatively less adequate parenting than did higher SES parents. However, little is known about the mechanisms through which lower SES is related to higher prevalence of child psychiatric disorders.

Because in our study chronic stress in the previous year was related to child psychiatric disorder, one SES related variable which would have been important to investigate was actual unemployment or the threat of possible unemployment. We carried out the population survey in the first half of 1983, a period in which Dutch economy and social securities rapidly went downhill, and which very likely adversely affected many families. Just as SES is a factor associated with psychiatric problems in children, economic factors may also be associated with child mental health. Graham (1978) even states "Perhaps the level of emotional and behavioral disturbance in its (i.e. developed countries) children is one criterion which governments eventually might come to regard as of considerable importance".

IX.5.1 Implications

A major implication which follows from our results is that studies within the field of developmental psychopathology always should test or control for age-, gender-, and SES-effects.

The difference in tendency between parents and clinicians to judge behavior problems in boys as more serious than in girls needs to be further investigated. Are girls better at hiding their symptoms from the clinician, do clinicians judge behavior problems in girls as less serious than in boys owing to qualitative differences in behavior, or are girls really less disturbed ? These are questions which need further clarification.

The gender differences found for aggressive behavior lead us to questions regarding etiologies of developmentally related phenomena.

Epidemiological studies have shown the importance of environmental influences on child psychopathology. However, epidemiological studies might also play a role in the investigation of the influence of genetic factors on disordered behavior, for instance by twin studies as was recently reported by Graham e.a. (1985).

The problem of the mechanisms through which SES exerts its influence on the development or maintenance of disturbed behavior in children is another field worth studying further.

Epidemiological methods can be used in order to arrive at a community diagnosis. Epidemiological surveys at regular intervals might reveal secular trends which can be associated with important cultural, economic and social changes in employment, in school system, or in education level.

The strength of this study (like Achenbach e.a.'s 1981) is that it is focussed on a broad age range, making age effects visible. However, the cross-sectional nature of this study did not permit inferences about the course the behavior of an individual child will follow. Longitudinal or sequential cohort designs (Schaie, 1965) are approaches to fill the knowledge-gap with respect to the development of disordered behavior and relationships between child and adult psychiatric disorders. The difficulties and disadvantages of longitudinal studies were discussed in chapter II. One of the problems was the rapid replacement of instruments because existing instruments did not meet the requirements. However, at this moment we have a number of good instruments at our disposal, making longitudinal studies more promising.

IX.6 The prevalence of psychiatric disorder - Implications

The number of 43 children (2.1%) in the general population sample (N=2076) referred to mental health agencies in the past year is in sharp contrast to the 26% moderately or severely disturbed 8- and 11-year-olds clinically judged to be in need of some kind of professional attention. Not all help can and must be given by child psychiatrists or clinical psychologists. Family- or school-doctors, pediatricians, teachers, school advisors or social workers familiar with child psychiatric problems can help parents and children with problems, or, in case the problem is too serious, pave the way for referral to a child psychiatric department. However, the prevalence rate of 26% moderately or severely disturbed children found in this study, warrants the continuation of efforts to get help for more children in the general population.

The prevalence rates we found were derived from a two-stage sampling method with clinical judgement as morbidity criterion in the second stage. It should be stressed that clinical judgement was a rather arbitrary one, probably explaining in part the difference in rates between our study (26% moderate or severe disorder) and the Isle of Wight study (12% moderate and severe disorder among 10-year-olds) (Rutter e.a., 1970). This stresses the need for more standardized assessment procedures. Studies that provide interrater reliability data for agreement between clinicians are carried

out by raters from the same institute. If diagnosticians in different countries could share the same training, the differences between studies might be reduced.

The number of children (116) intensively studied was too small to derive prevalence rates for specific disorders. An instrument like the Child Behavior Profile (Achenbach e.a, 1983) can provide readily obtainable measures on the level of more specific behavior disorders.

IX.7 Factors associated with psychiatric disorder

IX.7.1 The child

Those specific items that are found to be strongly associated with psychiatric disorder by different methods and by different studies can be regarded as very strong indicators of psychopathology. Seven CBCL-items for which referral status showed a large effect in ANCOVAs of behavior problems in our study, as well as in Achenbach e.a.'s (1981) study, were also among the best discriminating items between the clinically defined problem- and comparison-group in the second stage of the study. These items are: 8, Can't concentrate; 9, Obsessions; 13, Confused; 19, Demands attention; 25, Poor peer relations; 45, Nervous; 103, Unhappy, sad or depressed. These items can be regarded as important indicators of changes in the child's condition, for instance as a result of therapeutic action. Striking was the finding that for item 103, Unhappy, sad or depressed Achenbach e.a. (1981) found the strongest association with referral status, whereas in our study we found the second strongest association for this item. Achenbach e.a. concluded from their finding that the upsurge in concern for childhood depression was well justified. Our data again justified their conclusion.

IX.7.2 The family - Implications

Our data confirmed the findings by Richman e.a. (1982) and Rutter e.a. (1974, 1976) that a number of adverse family and environmental factors are associated with psychiatric disorder in the child (see table VII.13 and VII.14). Lack of warmth of the mother towards the child, conflicts between child and mother, poor marital relationships, as well as chronic stress and mental health problems in the mother, are significantly associated with child psychiatric disorder. Richman e.a.'s (1981) study showed the crucial role which disturbed family relationships play in the development of disorders in the child. The majority of 3-year-old nondisturbed children from families with disturbed relationships became disturbed by the time they were 8.

SUMMARY

In chapter I the history of child psychiatric epidemiology is described. The study is designed (1) to provide prevalence data on emotional and behavioral problems and on competencies of children in a general population sample, (2) to identify differences related to demographic variables, (3) to provide additional data on the validity of the Child Behavior Checklist (CBCL) - an instrument desgined to collect data on a wide variety of clinically relevant child behaviors in a broad age range, (4) to compare Dutch data with data from other surveys reported in the literature, especially with data from the study of Achenbach e.a. (1981), (5) to relate data obtained by the CBCL to other assessment procedures.

In chapter II the literature is reviewed. Special attention is paid to basic concepts and methods in child psychiatric epidemiology. Major limitations are the great differences and arbitrariness in defining psychiatric disorder and the lack of instruments tested in populations different from the ones for which the instruments were originally developed.

In chapter III the methods of the first, extensive, stage of the study are outlined. Parental reports were obtained with the CBCL in home interviews on 2076 children aged 4-16, randomly selected from the general population. A response rate of 85.1% was achieved. A normative sample was composed by excluding from the general population sample 43 children who were referred to a mental health agency in the year prior to the date of the interview. In addition, on 73% of children aged 4-11 from the start population, information from the teacher was obtained using the Teacher Report Form (TRF), an instrument analogous to the CBCL. Furthermore, parents of 1387 children referred to mental health agencies completed the CBCL at intake. Intraclass correlation coefficients were computed for test-retest, interrater, and interparent agreement. The ICC's obtained ranged from .69 to .99.

In chapter IV the results of the first stage are presented. The rate with which each CBCL behavior problem item was reported, is graphically depicted. ANCOVAs performed to assess the effects and interactions of age, gender, and SES in the general population sample for every behavior problem and social competence item of the CBCL, revealed numerous differences. There was a tendency for behavior problems to decline with age. Boys obtained only slightly higher mean total behavior problem scores than girls. More pronounced was the tendency for boys to score higher on socially disapproved behavior. For lower SES children, more behavior problems and fewer social competencies were reported. There was a tendency for lower SES parents to score their children higher on problems associated with the Externalizing syndrome. The percentage of children in the nonreferred sample for whom grade repetition was reported strongly increased after the transition from elementary to secondary school had taken place. The CBCL was found to discriminate between normal and referred children with an overall misclassification rate of 18.9% using cutoff scores corresponding to the 90th percentile of total behavior problem

scores and the 10th percentile of social competence scores in the nonreferred sample. Weighing items on the basis of discriminant functions did not improve the CBCL's discriminative power. Most items found to be strongly associated with referral status were also among the items showing large effects of clinical status in Achenbach e.a.'s study.

In chapter V the findings are discussed. Dutch parents tended to score their children higher on externalizing or undercontrolled behaviors than parents in other studies. However, the mean total behavior problem scores and the mean total social competence scores found in Achenbach e.a.'s study and in ours are strikingly similar. The data presented in this study are of importance to the Dutch situation, because no instruments designed to capture a wide variety of childhood behavior problems had been normed for the Dutch population before. The numerous age, gender and SES effects were discussed in light of their contribution to other data or theories.

In chapter VI the methods of the second, more clinically oriented, intensive stage are discussed. In this part of the study, 8- and 11-year-old children were selected for direct clinical assessment, if their scores on the CBCL and/or on the TRF were at or above the 85th percentile of the cumulative frequency distribution of each age. From the remaining 85% of each of the two populations, a normal group was randomly selected and clinically assessed. Of the 153 children selected, 116 (75.8%) children and their parents were interviewed. The assessment of the child carried out by 2 child psychiatrists, consisted of a structured child psychiatric interview, the WISC-R short form, fine motor tests, and drawing. Parental information on the child, as well as on a number of personal and family characteristics, was obtained by interviewers who conducted a semi-structured interview. Ratings on the severity of psychiatric disorder were made by the psychiatrists who conducted the child interviews directly after the child had been assessed. Overall ratings and diagnoses based on written information from the clinician and from the parent(s) were carried out by three child psychiatrists. Interrater and inter-interviewer reliabilities were satisfactory.

In chapter VII the results are presented. The Kendall correlation between CBCL and direct child assessment is .42, whereas the correlation between the TRF and direct child assessment is .28. The correlation between TRF and CBCL is .26. Prevalence rates for the whole population of 8- and 11-year-olds were estimated. Seven percent of the 8- and 11-year-olds were judged severely disordered, whereas for 26% the clinicians judged the child to be moderately or severely disordered. The sample size was too small to derive prevalence rates for specific diagnoses. ANOVAs and discriminant function analysis were performed to obtain those CBCL items that showed the best discrimination between children who were clinically judged disordered and children from the comparison group. Many items found to be good discriminators of clinical status in this stage of the study were also among the best discriminators of clinical status in the first stage of the study. A number of family and social factors were found to be associated with psychiatric disorder.

In chapter VIII the results of the second stage are discussed. The

implications of the low correlations between raters who saw children in different situations are emphasized. The weight given to reports from different sources varies with the kind of problem the child exhibits. It is concluded that more research is needed to investigate the relative value of different types of data for different conditions. The rather high prevalence rates found in our study compared with others is partly attributed to the arbitrary nature of clinical judgement. Family and social factors found in our study to be associated with psychiatric disorder support findings in other studies.

In chapter IX the study is evaluated and suggestions for further research are given. Attention is paid to the practical consequences of our findings especially regarding the use of the CBCL.

APPENDIX A

CHILD BEHAVIOR CHECKLIST FOR AGES 4-16

For office use only
ID #

CHILD'S NAME

SEX ☐ Boy ☐ Girl

AGE

RACE

TODAY'S DATE
Mo. ____ Day ____ Yr. ____

CHILD'S BIRTHDATE
Mo. ____ Day ____ Yr. ____

GRADE IN SCHOOL

PARENT'S TYPE OF WORK *(Please be specific—for example: auto mechanic, high school teacher, homemaker, laborer, lathe operator, shoe salesman, army sergeant, even if parent does not live with child.)*

FATHER'S TYPE OF WORK: _____

MOTHER'S TYPE OF WORK: _____

THIS FORM FILLED OUT BY:
☐ Mother
☐ Father
☐ Other *(Specify)*:

I. Please list the sports your child most likes to take part in. For example: swimming, baseball, skating, skate boarding, bike riding, fishing, etc.
☐ None

	Compared to other children of the same age, about how much time does he/she spend in each?				Compared to other children of the same age, how well does he/she do each one?			
	Don't Know	Less Than Average	Average	More Than Average	Don't Know	Below Average	Average	Above Average
a. _____	☐	☐	☐	☐	☐	☐	☐	☐
b. _____	☐	☐	☐	☐	☐	☐	☐	☐
c. _____	☐	☐	☐	☐	☐	☐	☐	☐

II. Please list your child's favorite hobbies, activities, and games, other than sports. For example: stamps, dolls, books, piano, crafts, singing, etc. (Do not include T.V.)
☐ None

	Compared to other children of the same age, about how much time does he/she spend in each?				Compared to other children of the same age, how well does he/she do each one?			
	Don't Know	Less Than Average	Average	More Than Average	Don't Know	Below Average	Average	Above Average
a. _____	☐	☐	☐	☐	☐	☐	☐	☐
b. _____	☐	☐	☐	☐	☐	☐	☐	☐
c. _____	☐	☐	☐	☐	☐	☐	☐	☐

III. Please list any organizations, clubs, teams, or groups your child belongs to.
☐ None

	Compared to other children of the same age, how active is he/she in each?			
	Don't Know	Less Active	Average	More Active
a. _____	☐	☐	☐	☐
b. _____	☐	☐	☐	☐
c. _____	☐	☐	☐	☐

IV. Please list any jobs or chores your child has. For example: paper route, babysitting, making bed, etc.
☐ None

	Compared to other children of the same age, how well does he/she carry them out?			
	Don't Know	Below Average	Average	Above Average
a. _____	☐	☐	☐	☐
b. _____	☐	☐	☐	☐
c. _____	☐	☐	☐	☐

3-81 Edition

V. 1. About how many close friends does your child have? ☐ None ☐ 1 ☐ 2 or 3 ☐ 4 or more

2. About how many times a week does your child do things with them? ☐ less than 1 ☐ 1 or 2 ☐ 3 or more

VI. Compared to other children of his/her age, how well does your child:

	Worse	About the same	Better
a. Get along with his/her brothers & sisters?	☐	☐	☐
b. Get along with other children?	☐	☐	☐
c. Behave with his/her parents?	☐	☐	☐
d. Play and work by himself/herself?	☐	☐	☐

VII. 1. Current school performance—for children aged 6 and older:

☐ Does not go to school

	Failing	Below average	Average	Above average
a. Reading or English	☐	☐	☐	☐
b. Writing	☐	☐	☐	☐
c. Arithmetic or Math	☐	☐	☐	☐
d. Spelling	☐	☐	☐	☐
Other academic subjects—for example: history, science, foreign language, geography. e. _____	☐	☐	☐	☐
f. _____	☐	☐	☐	☐
g. _____	☐	☐	☐	☐

2. Is your child in a special class?

☐ No ☐ Yes—what kind?

3. Has your child ever repeated a grade?

☐ No ☐ Yes—grade and reason

4. Has your child had any academic or other problems in school?

☐ No ☐ Yes—please describe

When did these problems start?

Have these problems ended?

☐ No ☐ Yes—when?

VIII. Below is a list of items that describe children. For each item that describes your child **now or within the past 6 months**, please circle the **2** if the item is **very true or often true** of your child. Circle the **1** if the item is **somewhat or sometimes true** of your child. If the item is **not true** of your child, circle the **0**. Please answer all items as well as you can, even if some do not seem to apply to your child.

0 = Not True (as far as you know) **1 = Somewhat or Sometimes True** **2 = Very True or Often True**

0 1 2 1. Acts too young for his/her age
0 1 2 2. Allergy (describe): _____
_____
0 1 2 3. Argues a lot
0 1 2 4. Asthma
0 1 2 5. Behaves like opposite sex
0 1 2 6. Bowel movements outside toilet
0 1 2 7. Bragging, boasting
0 1 2 8. Can't concentrate, can't pay attention for long
0 1 2 9. Can't get his/her mind off certain thoughts; obsessions (describe): _____
_____
0 1 2 10. Can't sit still, restless, or hyperactive
0 1 2 11. Clings to adults or too dependent
0 1 2 12. Complains of loneliness
0 1 2 13. Confused or seems to be in a fog
0 1 2 14. Cries a lot
0 1 2 15. Cruel to animals
0 1 2 16. Cruelty, bullying, or meanness to others
0 1 2 17. Day-dreams or gets lost in his/her thoughts
0 1 2 18. Deliberately harms self or attempts suicide
0 1 2 19. Demands a lot of attention
0 1 2 20. Destroys his/her own things
0 1 2 21. Destroys things belonging to his/her family or other children
0 1 2 22. Disobedient at home
0 1 2 23. Disobedient at school
0 1 2 24. Doesn't eat well
0 1 2 25. Doesn't get along with other children
0 1 2 26. Doesn't seem to feel guilty after misbehaving
0 1 2 27. Easily jealous
0 1 2 28. Eats or drinks things that are not food (describe): _____
_____
0 1 2 29. Fears certain animals, situations, or places, other than school (describe): _____
_____
0 1 2 30. Fears going to school
0 1 2 31. Fears he/she might think or do something bad
0 1 2 32. Feels he/she has to be perfect
0 1 2 33. Feels or complains that no one loves him/her
0 1 2 34. Feels others are out to get him/her
0 1 2 35. Feels worthless or inferior
0 1 2 36. Gets hurt a lot, accident-prone
0 1 2 37. Gets in many fights
0 1 2 38. Gets teased a lot
0 1 2 39. Hangs around with children who get in trouble
0 1 2 40. Hears things that aren't there (describe):
_____
_____
0 1 2 41. Impulsive or acts without thinking
0 1 2 42. Likes to be alone
0 1 2 43. Lying or cheating
0 1 2 44. Bites fingernails
0 1 2 45. Nervous, highstrung, or tense
0 1 2 46. Nervous movements or twitching (describe):
_____
_____
0 1 2 47. Nightmares
0 1 2 48. Not liked by other children
0 1 2 49. Constipated, doesn't move bowels
0 1 2 50. Too fearful or anxious
0 1 2 51. Feels dizzy
0 1 2 52. Feels too guilty
0 1 2 53. Overeating
0 1 2 54. Overtired
0 1 2 55. Overweight
56. Physical problems without known medical cause:
0 1 2 a. Aches or pains
0 1 2 b. Headaches
0 1 2 c. Nausea, feels sick
0 1 2 d. Problems with eyes (describe):
_____
0 1 2 e. Rashes or other skin problems
0 1 2 f. Stomachaches or cramps
0 1 2 g. Vomiting, throwing up
0 1 2 h. Other (describe): _____
_____

Please see other side

0 = Not True (as far as you know) 1 = Somewhat or Sometimes True 2 = Very True or Often True

0 1 2 57. Physically attacks people
0 1 2 58. Picks nose, skin, or other parts of body (describe): _____

_____ 80

0 1 2 59. Plays with own sex parts in public 15
0 1 2 60. Plays with own sex parts too much

0 1 2 61. Poor school work
0 1 2 62. Poorly coordinated or clumsy

0 1 2 63. Prefers playing with older children 20
0 1 2 64. Prefers playing with younger children

0 1 2 65. Refuses to talk
0 1 2 66. Repeats certain acts over and over; compulsions (describe): _____

_____

0 1 2 67. Runs away from home
0 1 2 68. Screams a lot 25

0 1 2 69. Secretive, keeps things to self
0 1 2 70. Sees things that aren't there (describe):

_____

_____

0 1 2 71. Self-conscious or easily embarrassed
0 1 2 72. Sets fires

0 1 2 73. Sexual problems (describe):

_____

_____ 30

0 1 2 74. Showing off or clowning

0 1 2 75. Shy or timid
0 1 2 76. Sleeps less than most children

0 1 2 77. Sleeps more than most children during day and/or night (describe): _____

_____

0 1 2 78. Smears or plays with bowel movements 35

0 1 2 79. Speech problem (describe): _____

_____

0 1 2 80. Stares blankly

0 1 2 81. Steals at home
0 1 2 82. Steals outside the home

0 1 2 83. Stores up things he/she doesn't need (describe):

_____ 40

0 1 2 84. Strange behavior (describe): _____

_____

0 1 2 85. Strange ideas (describe):

_____

0 1 2 86. Stubborn, sullen, or irritable

0 1 2 87. Sudden changes in mood or feelings
0 1 2 88. Sulks a lot 45

0 1 2 89. Suspicious
0 1 2 90. Swearing or obscene language

0 1 2 91. Talks about killing self
0 1 2 92. Talks or walks in sleep (describe):

_____

0 1 2 93. Talks too much 50
0 1 2 94. Teases a lot

0 1 2 95. Temper tantrums or hot temper
0 1 2 96. Thinks about sex too much

0 1 2 97. Threatens people
0 1 2 98. Thumb-sucking 55

0 1 2 99. Too concerned with neatness or cleanliness
0 1 2 100. Trouble sleeping (describe):

_____

0 1 2 101. Truancy, skips school
0 1 2 102. Underactive, slow moving, or lacks energy

0 1 2 103. Unhappy, sad, or depressed 60
0 1 2 104. Unusually loud

0 1 2 105. Uses alcohol or drugs (describe):

_____

0 1 2 106. Vandalism

0 1 2 107. Wets self during the day
0 1 2 108. Wets the bed 65

0 1 2 109. Whining
0 1 2 110. Wishes to be of opposite sex

0 1 2 111. Withdrawn, doesn't get involved with others
0 1 2 112. Worrying

113. Please write in any problems your child has that were not listed above:

0 1 2 _____ 70

0 1 2 _____

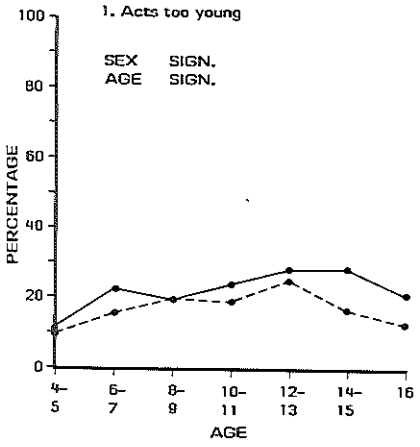
0 1 2 _____

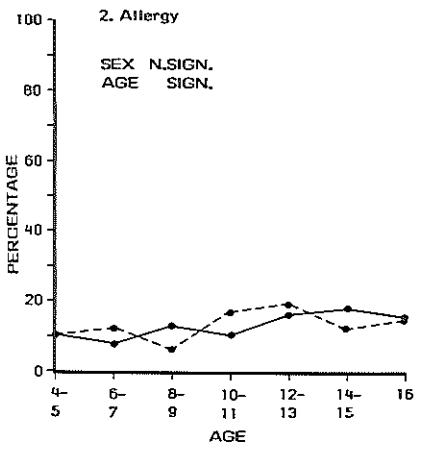
PLEASE BE SURE YOU HAVE ANSWERED ALL ITEMS. PAGE 4 UNDERLINE ANY YOU ARE CONCERNED ABOUT.

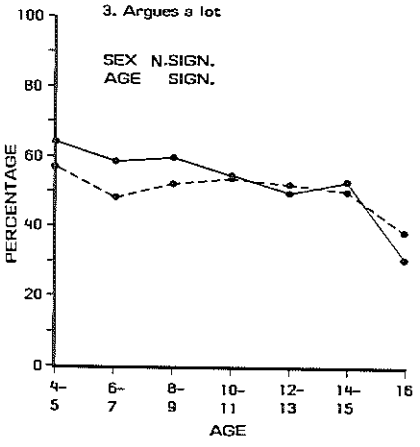
APPENDIX B

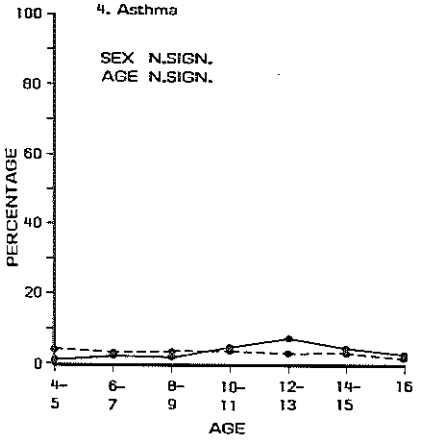
Figures 1-112 - Percentage of children in the general population sample (N=2076) of each gender for whom each behavior problem was reported by the parents.
Scores of 1 and 2 for each item are combined to obtain the percentage for whom the problem was reported.

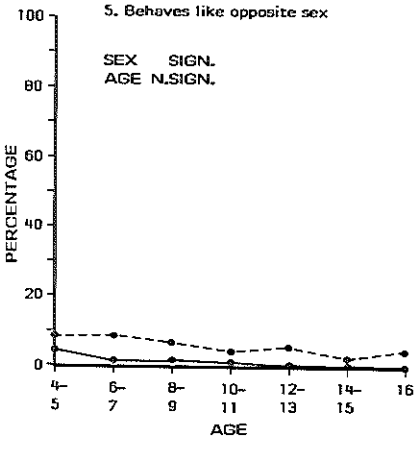
------- Girls
——— Boys
Sign. = Significant ($P<0.01$) effect in ANCOVAs of behavior problems
N.Sign. = Non-significant ($P\geq 0.01$) effect in ANCOVAs of behavior problems

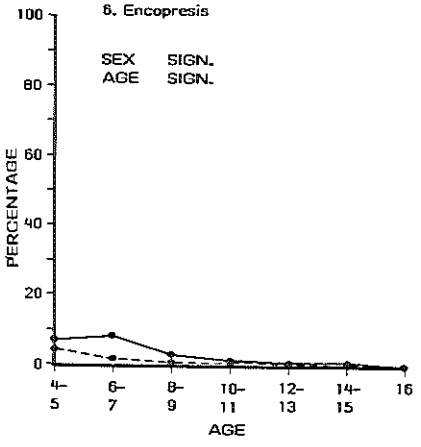

1. Acts too young
SEX SIGN.
AGE SIGN.
PERCENTAGE
AGE

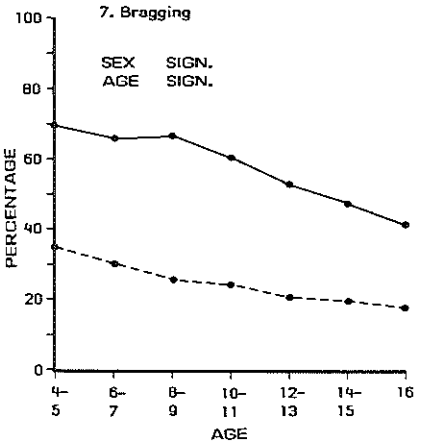

2. Allergy
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
AGE

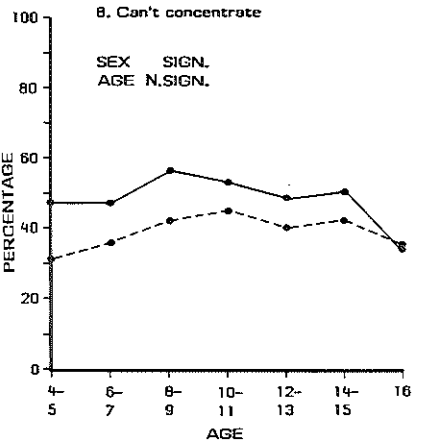

3. Argues a lot
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
AGE

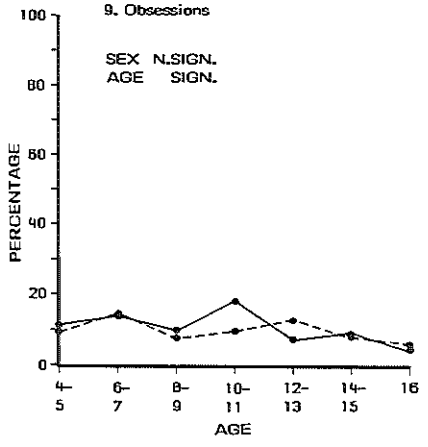

4. Asthma
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

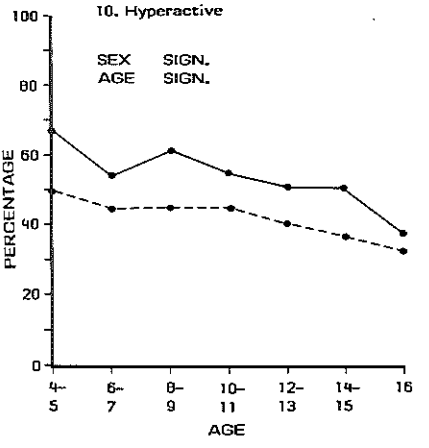

5. Behaves like opposite sex
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

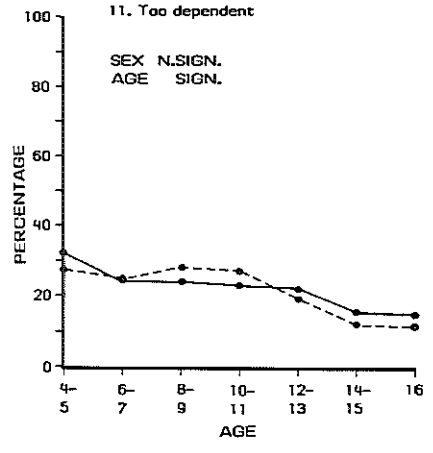

6. Encopresis
SEX SIGN.
AGE SIGN.
PERCENTAGE
AGE

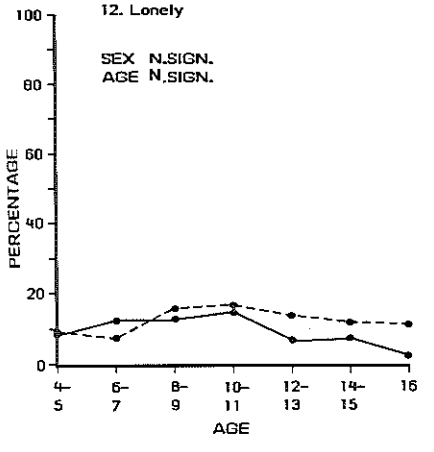

7. Bragging
SEX SIGN.
AGE SIGN.
PERCENTAGE
AGE

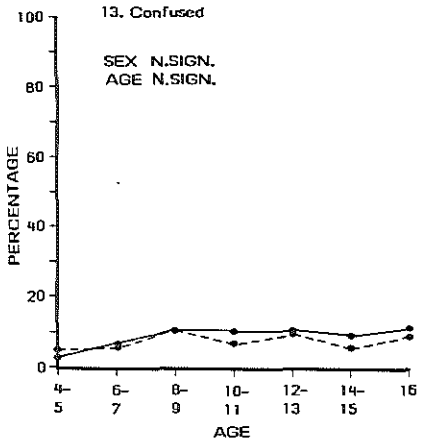

8. Can't concentrate
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

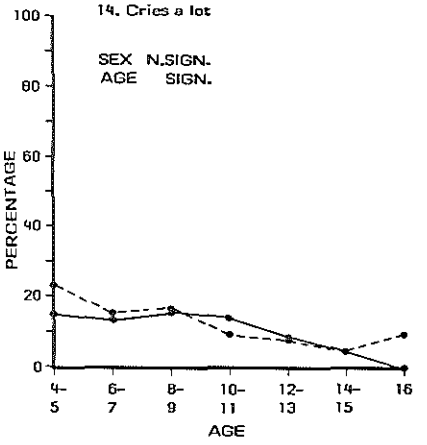

9. Obsessions
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
AGE

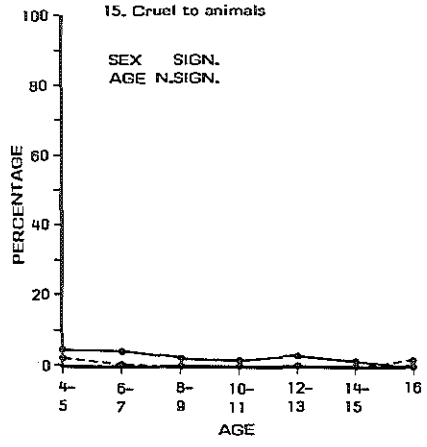

10. Hyperactive
SEX SIGN.
AGE SIGN.
PERCENTAGE
AGE

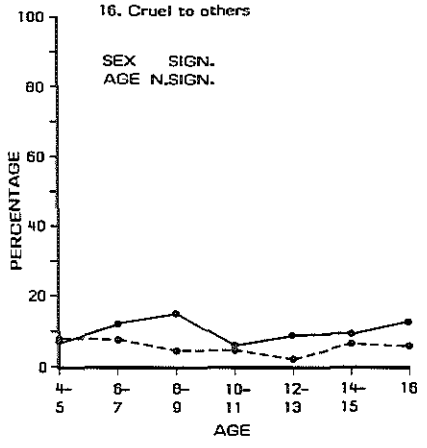

11. Too dependent
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
AGE

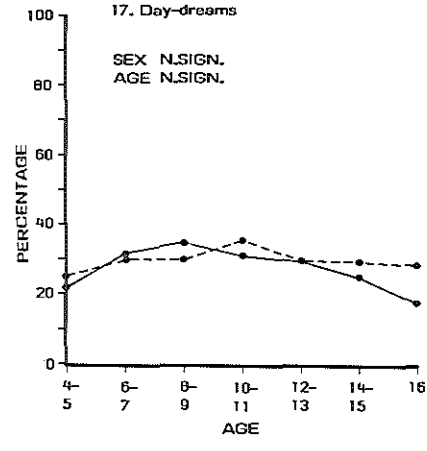

12. Lonely
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

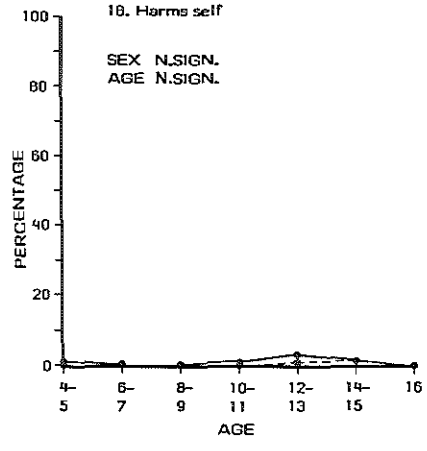

13. Confused
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

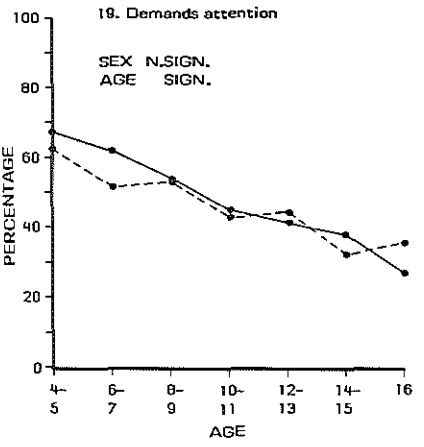

14. Cries a lot
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
AGE

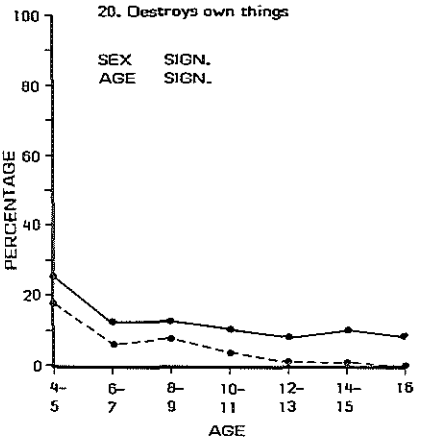

15. Cruel to animals
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

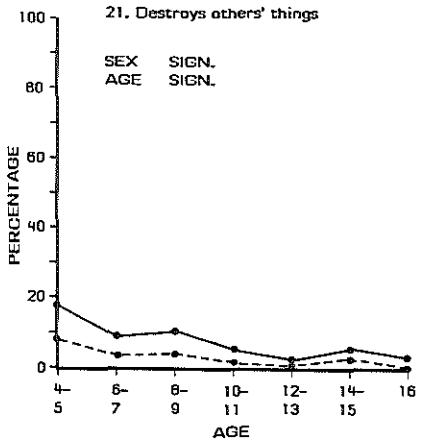

16. Cruel to others
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

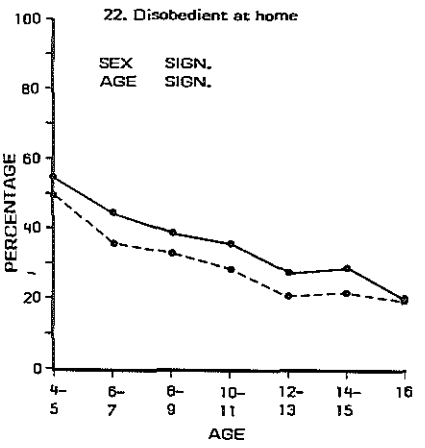

17. Day-dreams
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

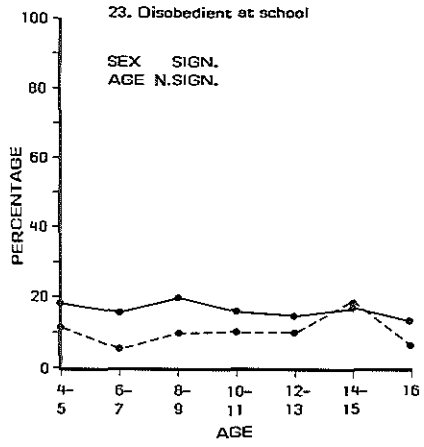

18. Harms self
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

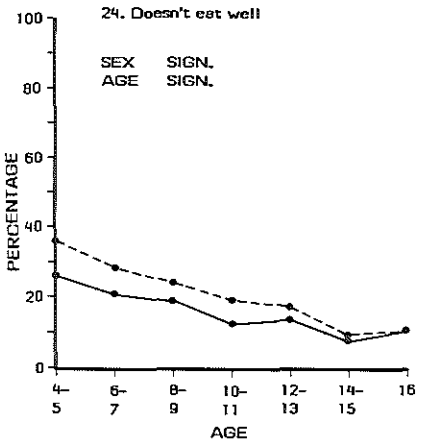

19. Demands attention
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

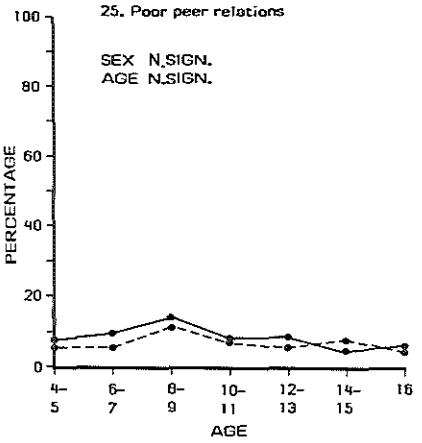

20. Destroys own things
SEX SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

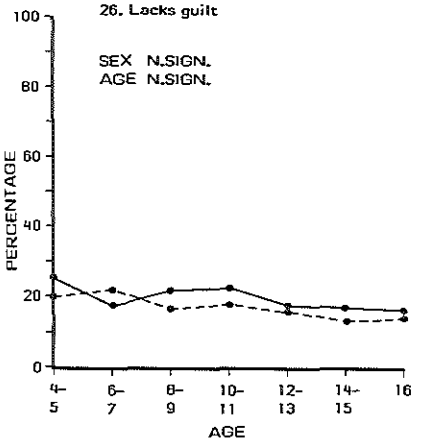

21. Destroys others' things
SEX SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

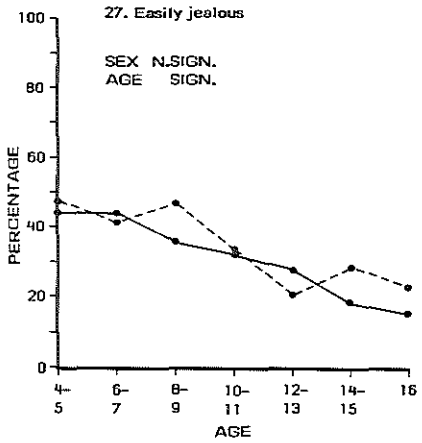

22. Disobedient at home
SEX SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

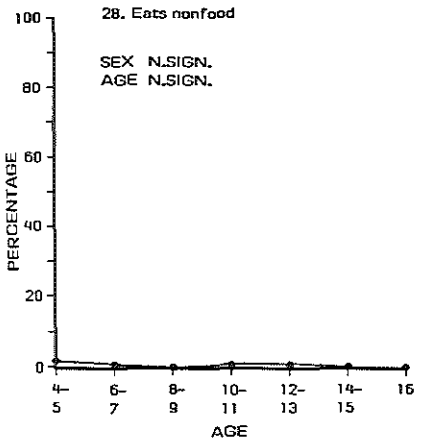

23. Disobedient at school
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

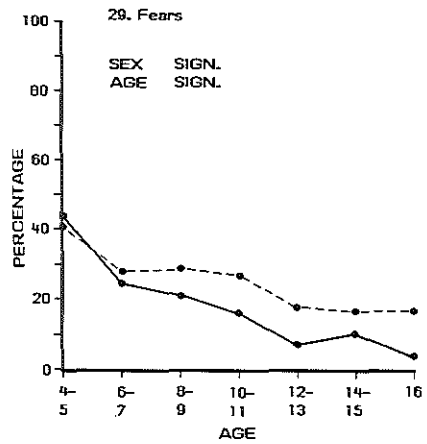

24. Doesn't eat well
SEX SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

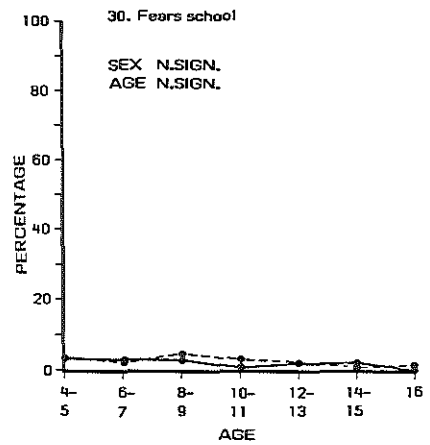

25. Poor peer relations
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

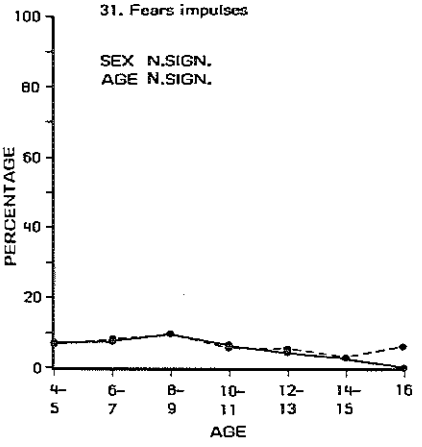

26. Lacks guilt
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

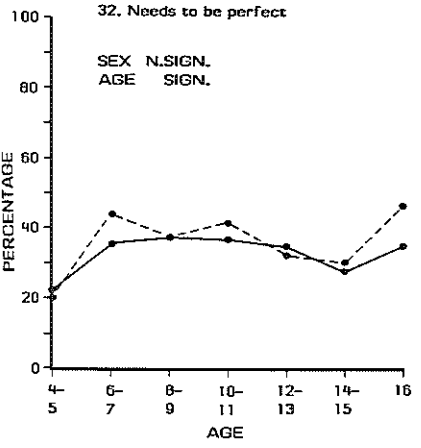

27. Easily jealous
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

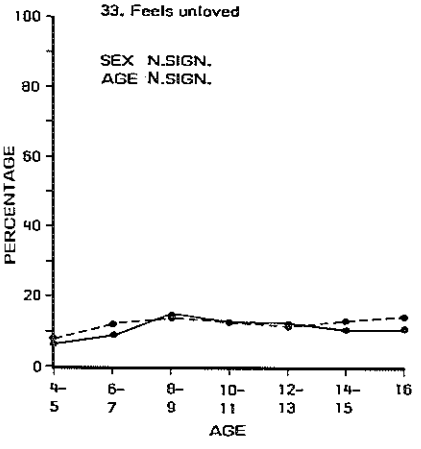

28. Eats nonfood
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

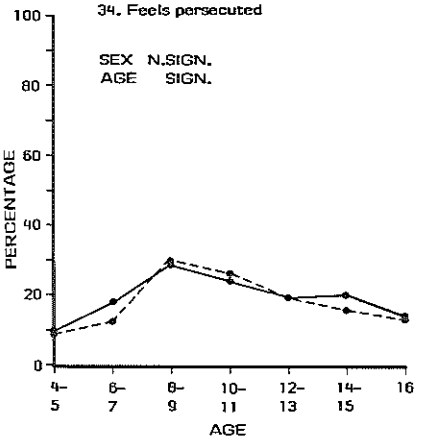

29. Fears
SEX SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

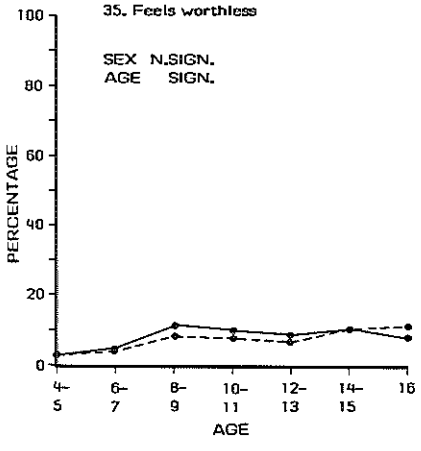

30. Fears school
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

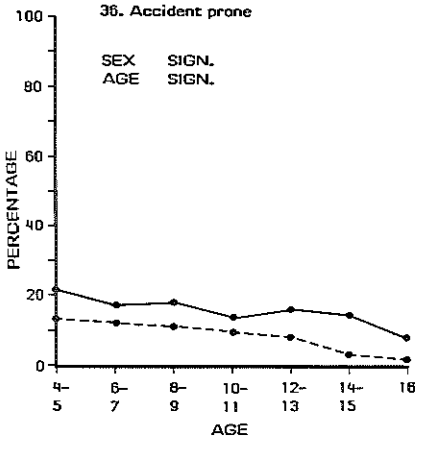

31. Fears impulses
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

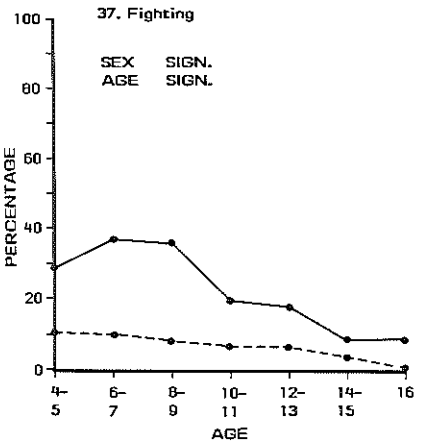

32. Needs to be perfect
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

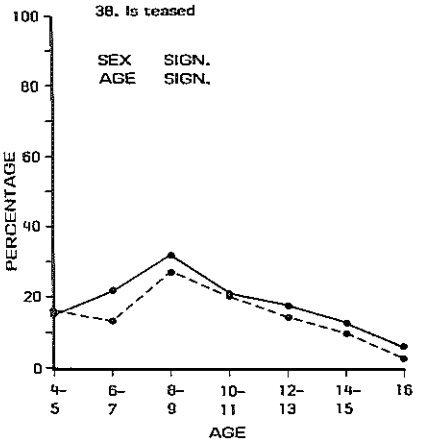

33. Feels unloved
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

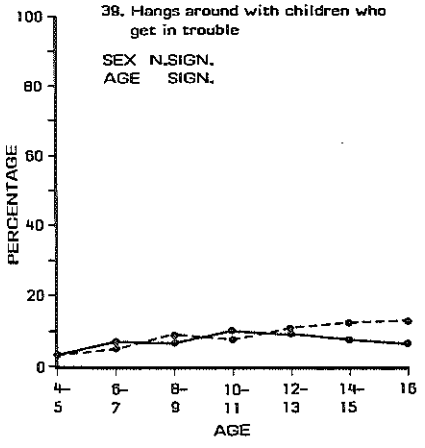

34. Feels persecuted
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

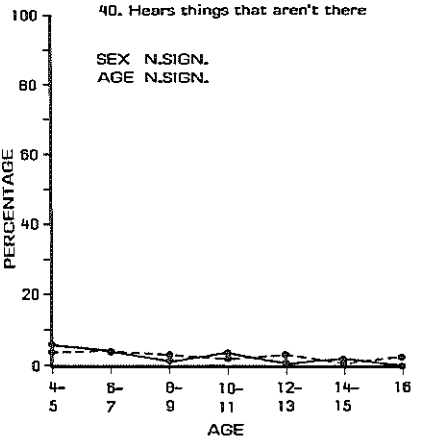

35. Feels worthless
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

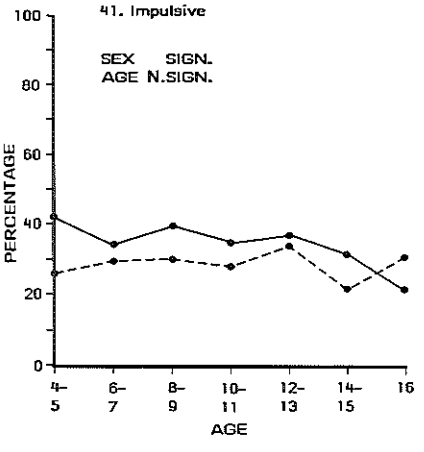

36. Accident prone
SEX SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

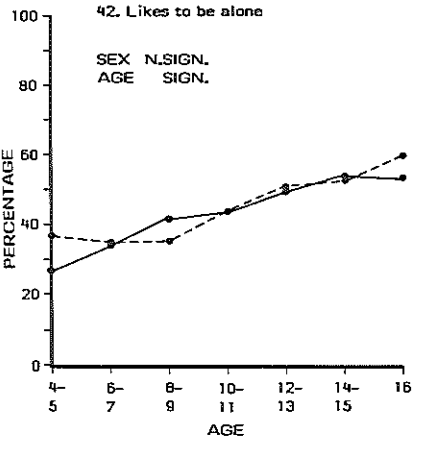

37. Fighting
SEX SIGN.
AGE SIGN.
PERCENTAGE
AGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16

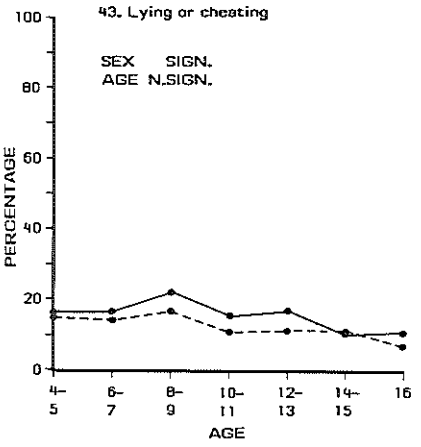

38. Is teased
SEX SIGN.
AGE SIGN.
PERCENTAGE
AGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16

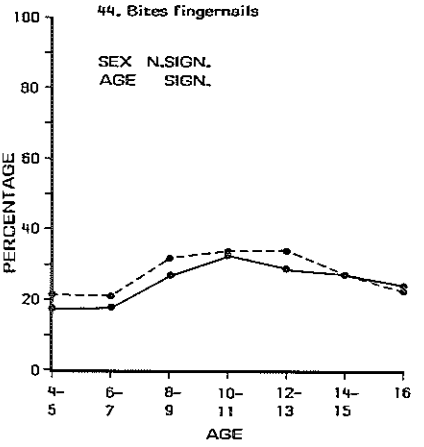

39. Hangs around with children who get in trouble
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
AGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16

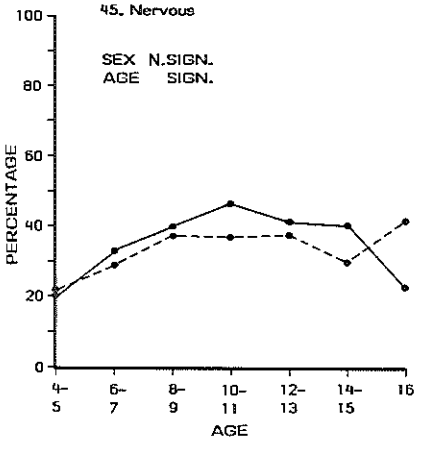

40. Hears things that aren't there
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
AGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16

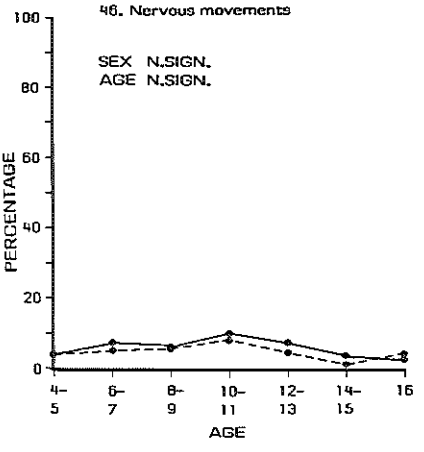

41. Impulsive
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
AGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16

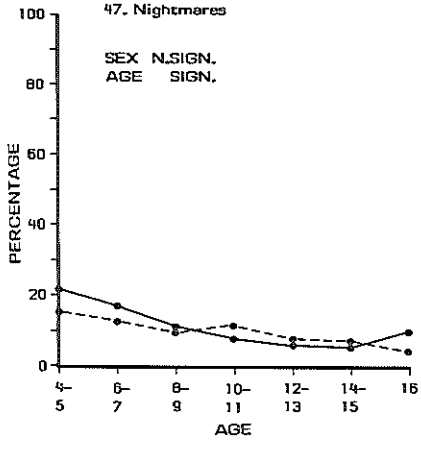

42. Likes to be alone
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
AGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16

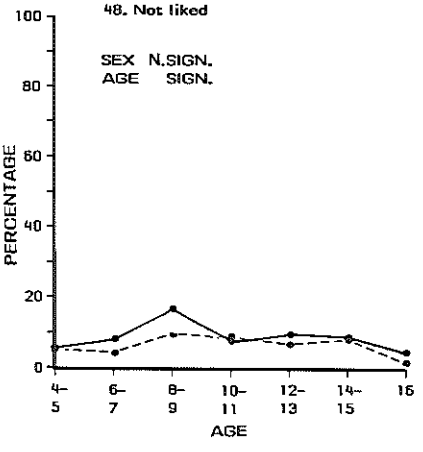

43. Lying or cheating
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

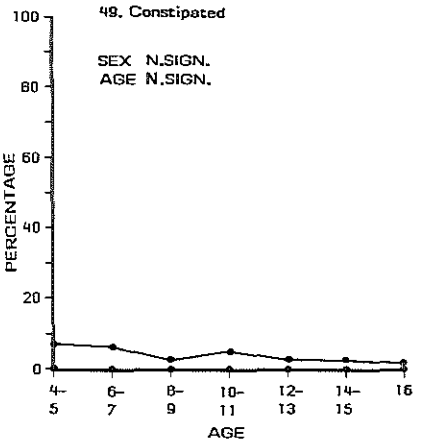

44. Bites fingernails
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

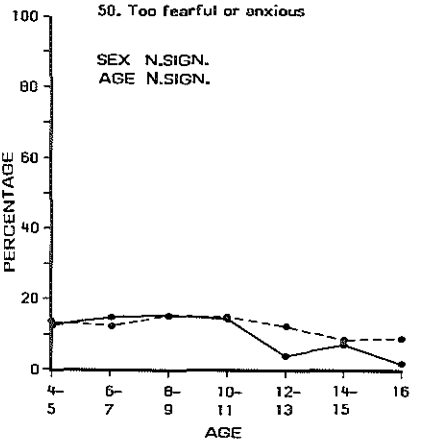

45. Nervous
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

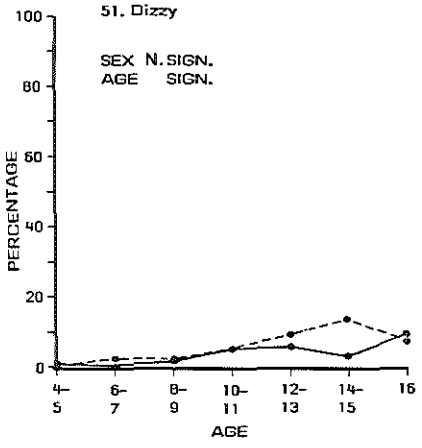

46. Nervous movements
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

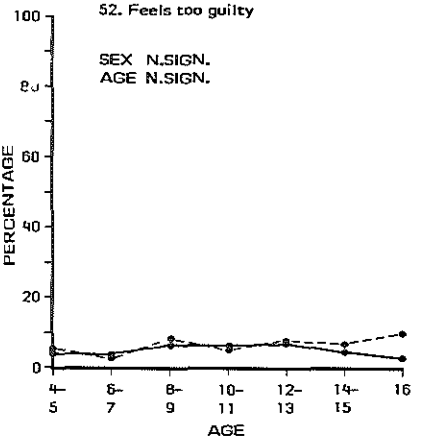

47. Nightmares
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

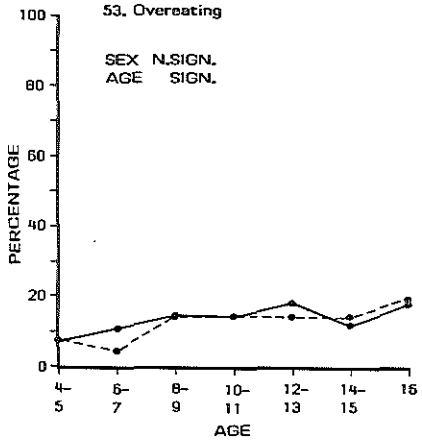

48. Not liked
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

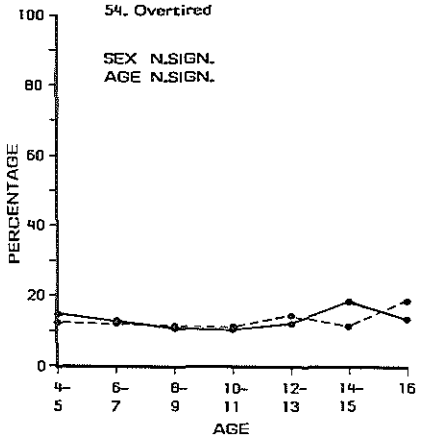

49. Constipated
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

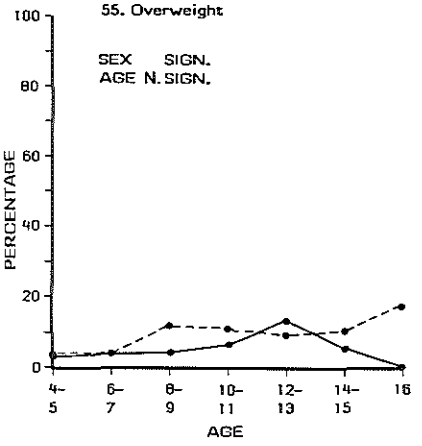

50. Too fearful or anxious
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

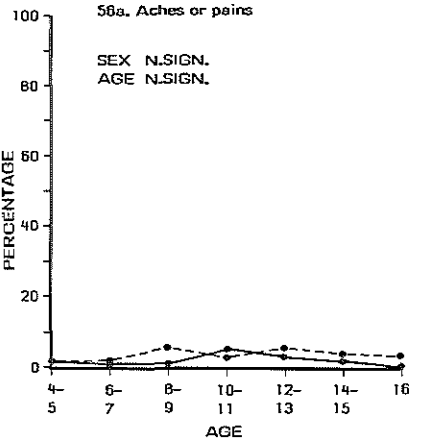

51. Dizzy
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

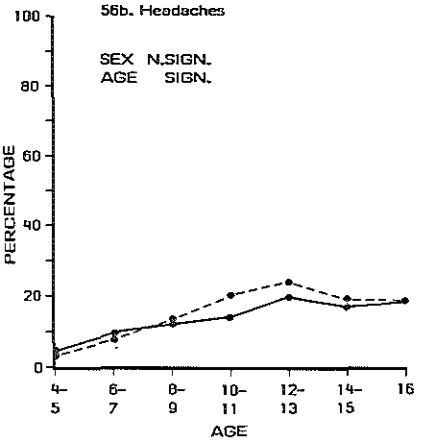

52. Feels too guilty
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

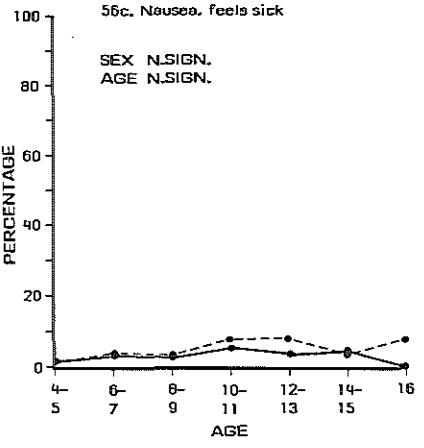

53. Overeating
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

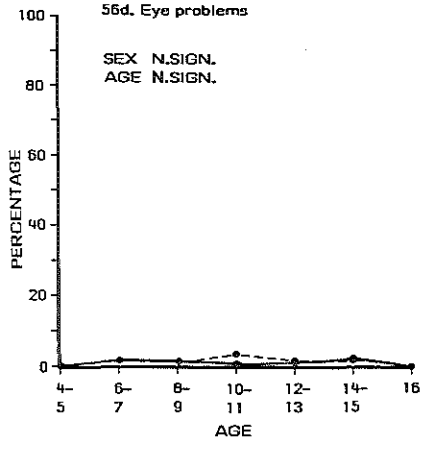

54. Overtired
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

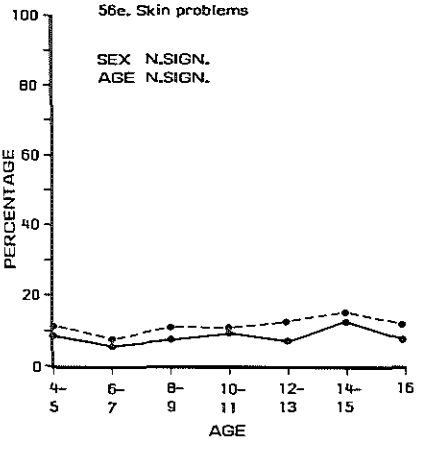

55. Overweight
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

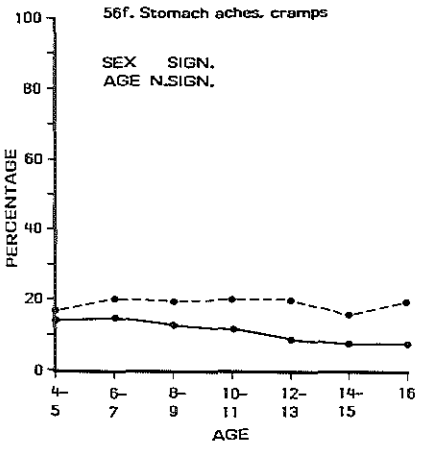

56a. Aches or pains
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

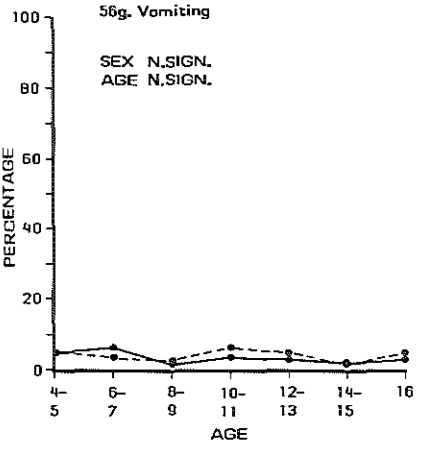

56b. Headaches
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

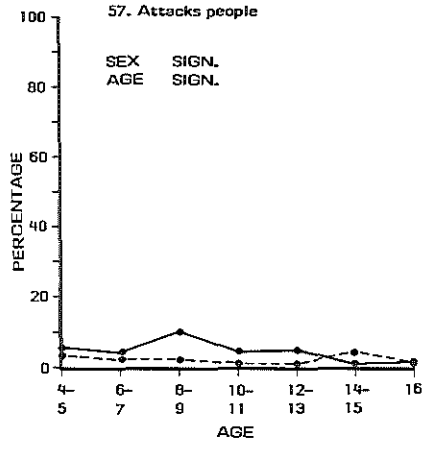

56c. Nausea, feels sick
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

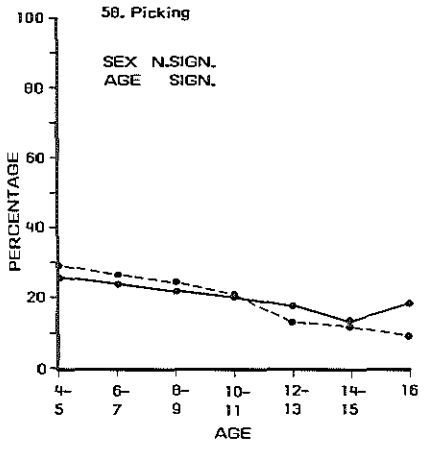

56d. Eye problems
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

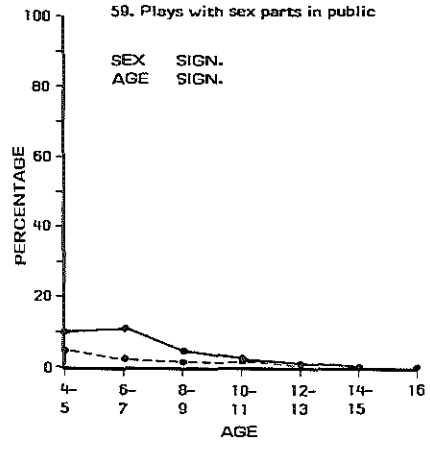

56e. Skin problems
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

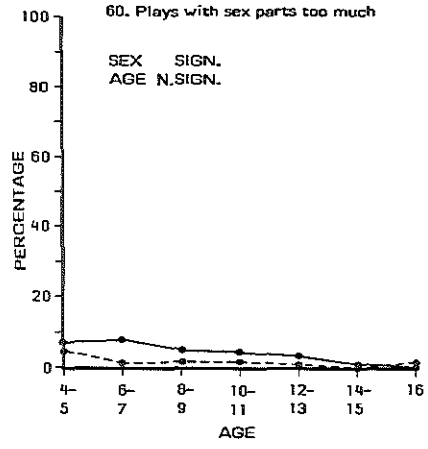

56f. Stomach aches. cramps
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

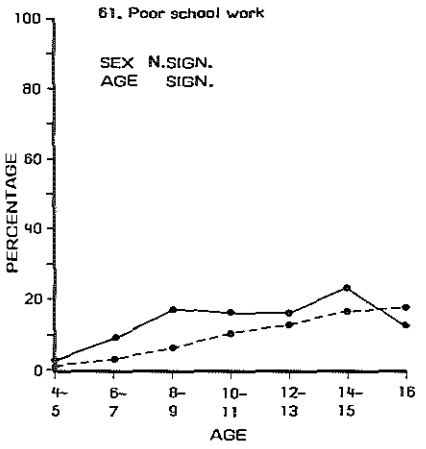

56g. Vomiting
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

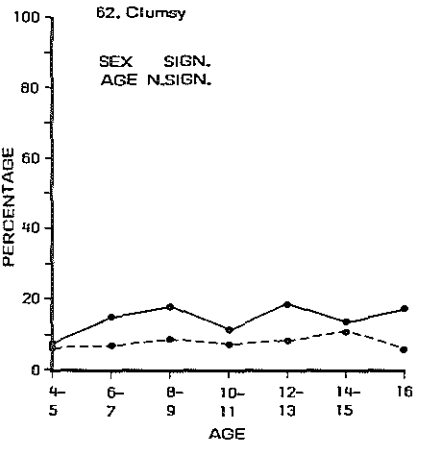

57. Attacks people
SEX SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

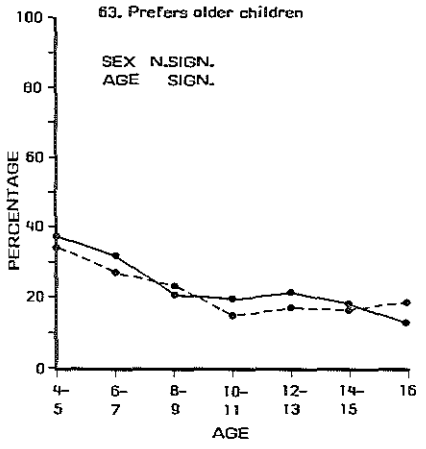

58. Picking
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

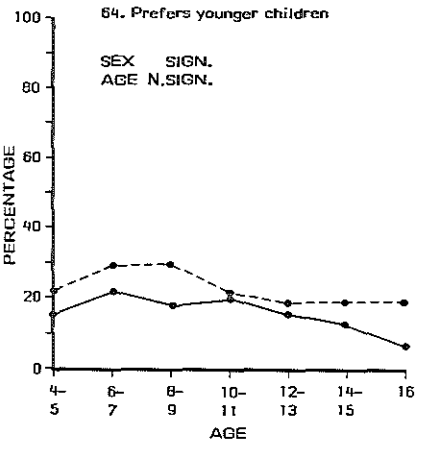

59. Plays with sex parts in public
SEX SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

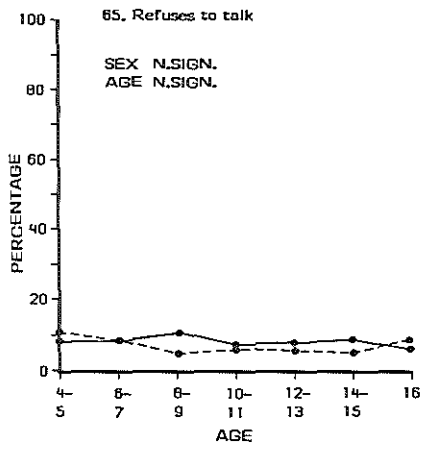

60. Plays with sex parts too much
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

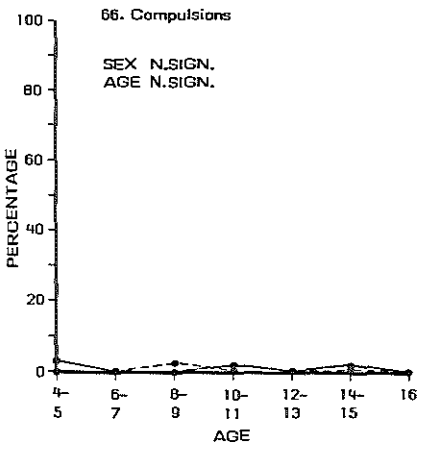

61. Poor school work
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

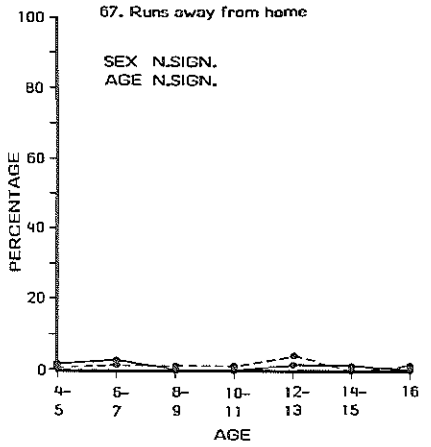

62. Clumsy
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

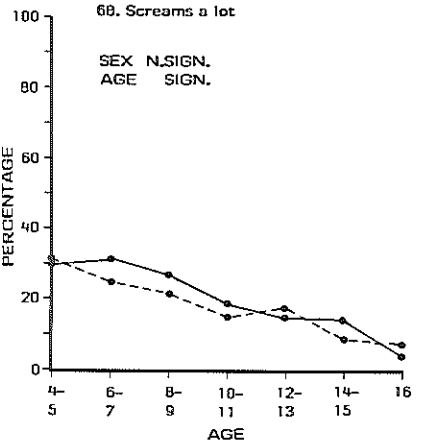

63. Prefers older children
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

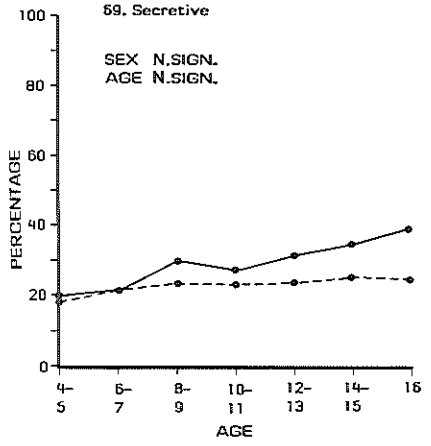

64. Prefers younger children
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

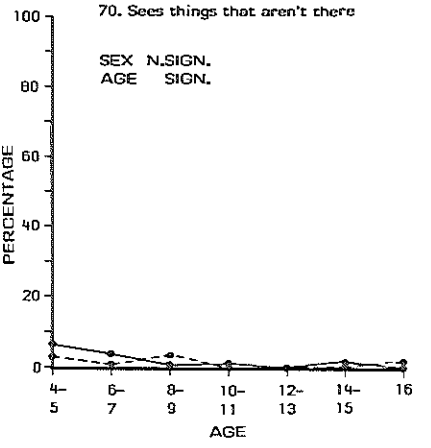

65. Refuses to talk
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

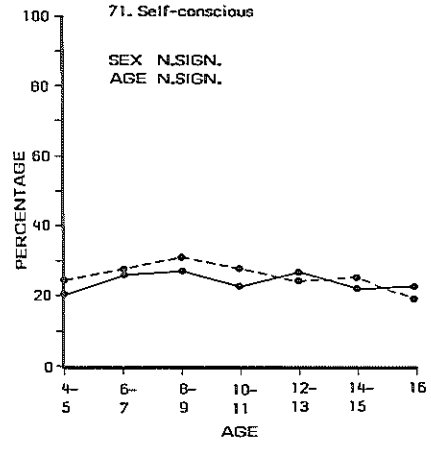

66. Compulsions
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

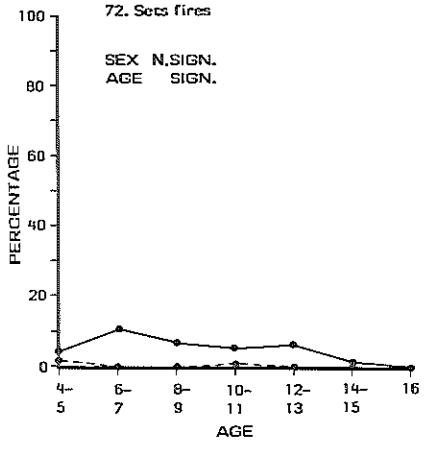

67. Runs away from home
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

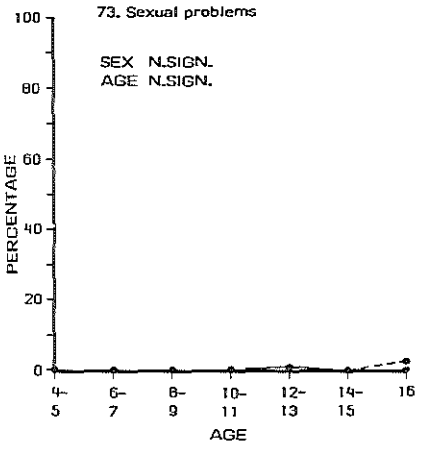

68. Screams a lot
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
AGE

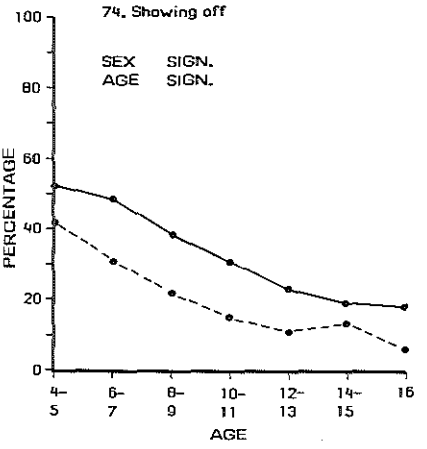

69. Secretive
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

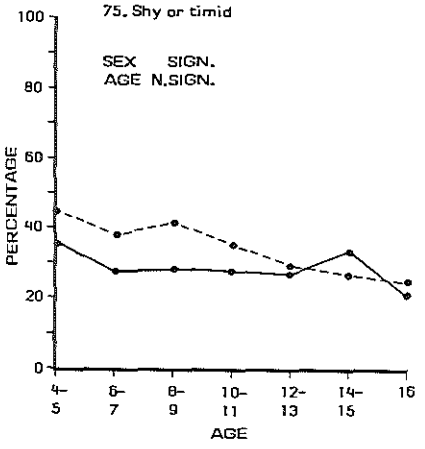

70. Sees things that aren't there
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
AGE

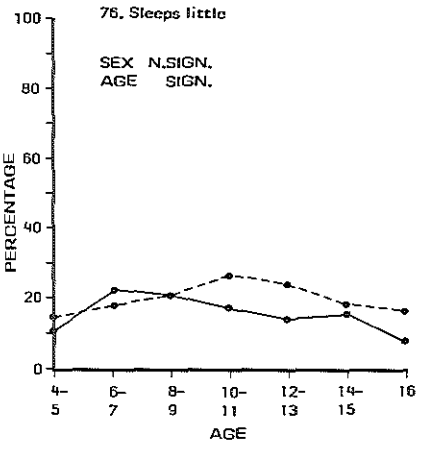

71. Self-conscious
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

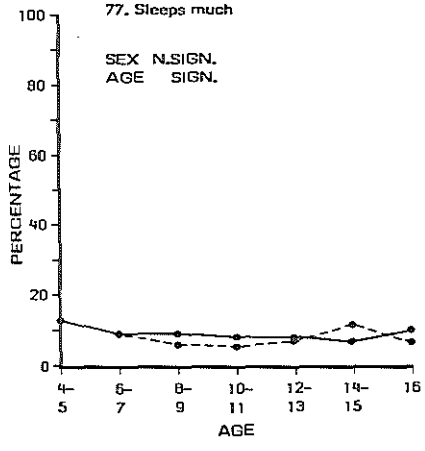

72. Sets fires
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
AGE

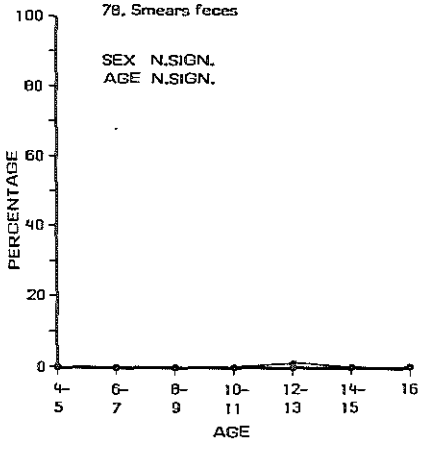

73. Sexual problems
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

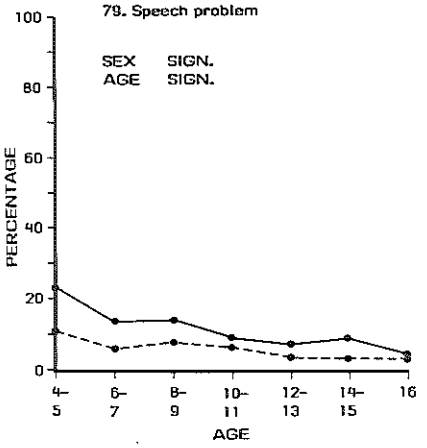

74. Showing off
SEX SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

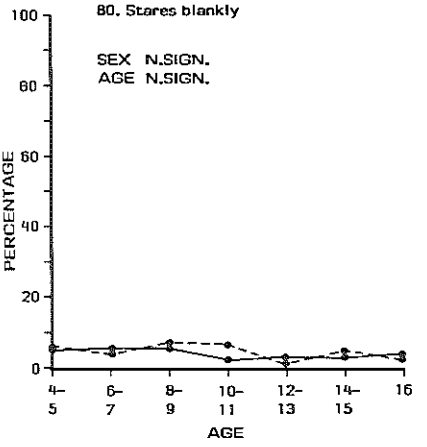

75. Shy or timid
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

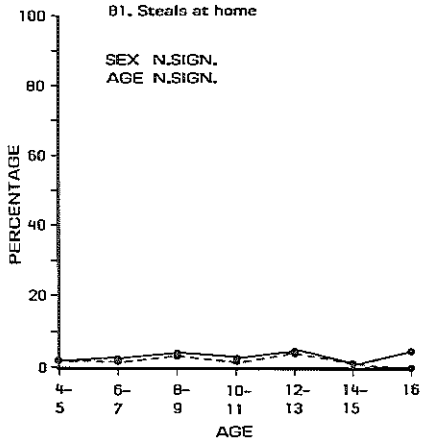

76. Sleeps little
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

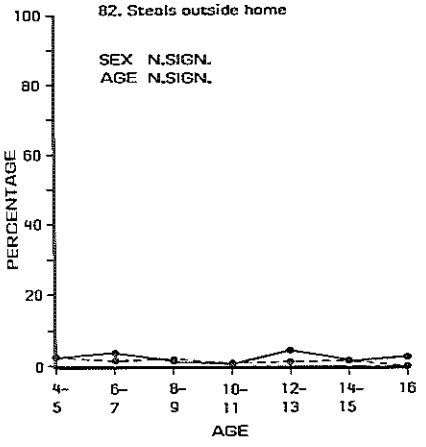

77. Sleeps much
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

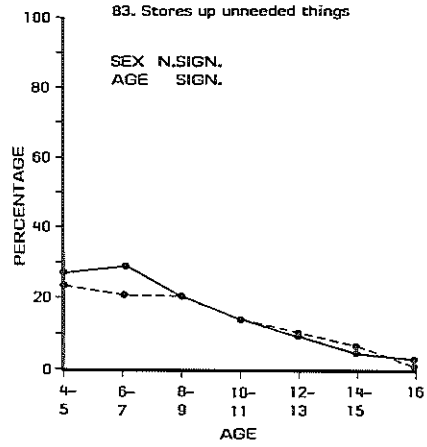

78. Smears feces
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

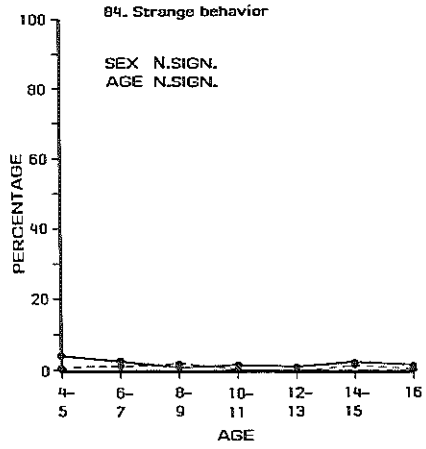

79. Speech problem
SEX SIGN.
AGE SIGN.
PERCENTAGE
AGE

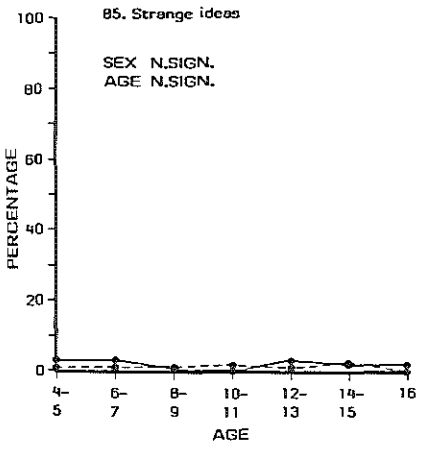

80. Stares blankly
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

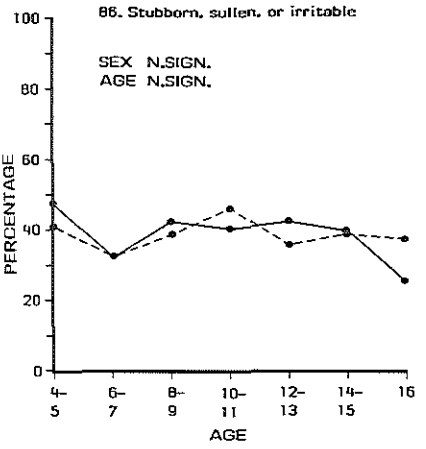

81. Steals at home
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

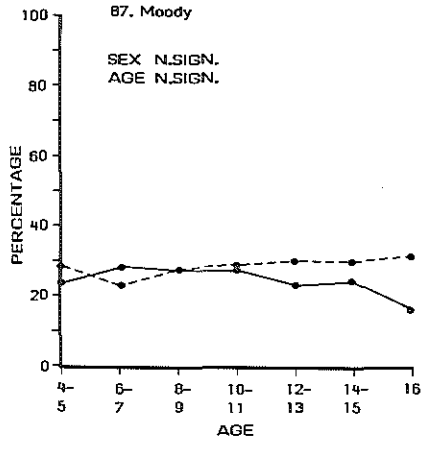

82. Steals outside home
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

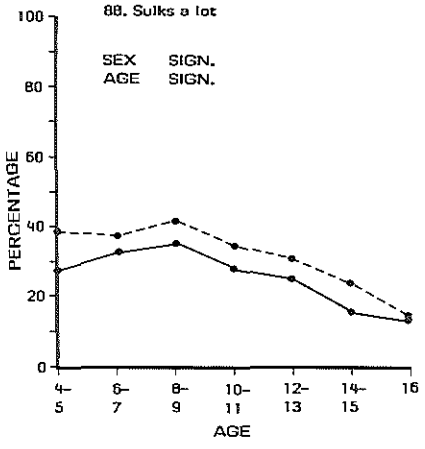

83. Stores up unneeded things
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
AGE

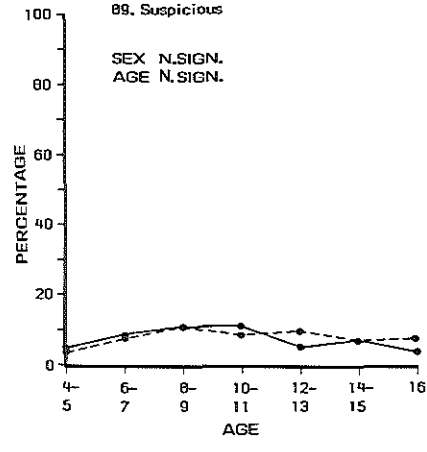

84. Strange behavior
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

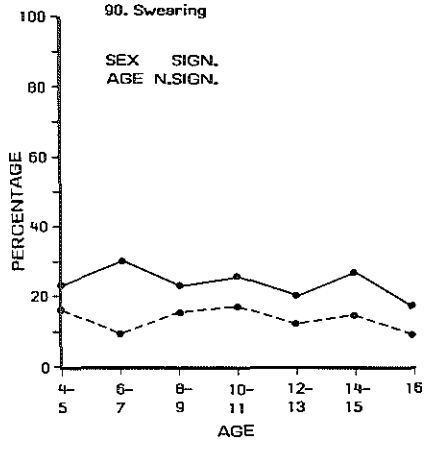

85. Strange ideas
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

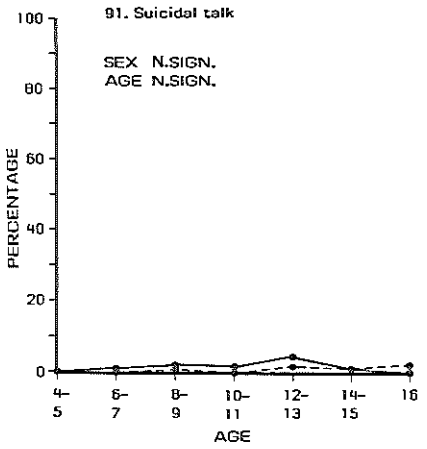

86. Stubborn, sullen, or irritable
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

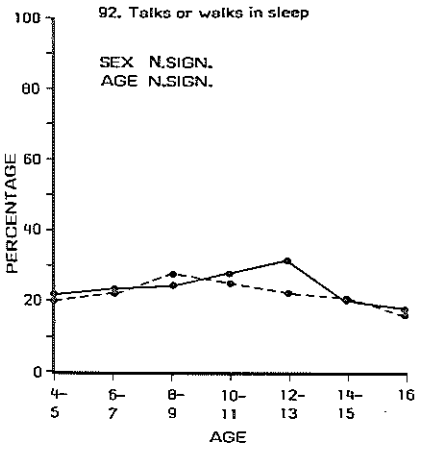

87. Moody
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

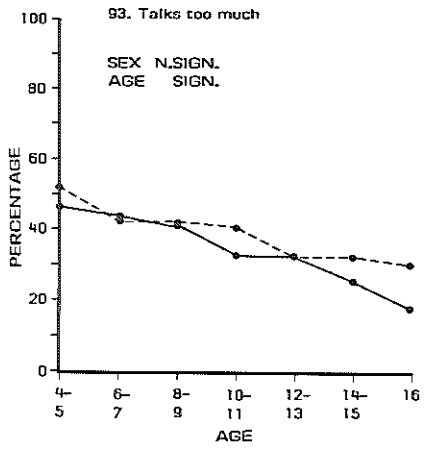

88. Sulks a lot
SEX SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

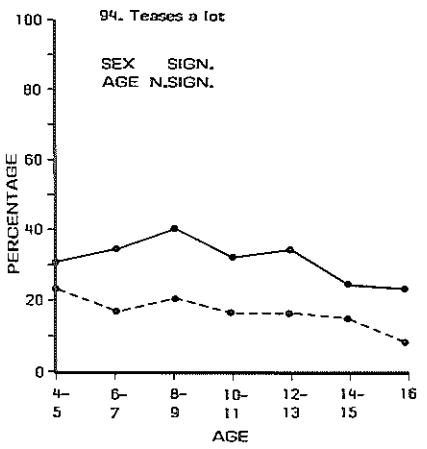

89. Suspicious
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

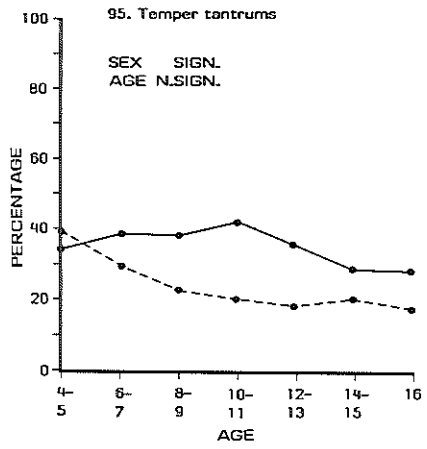

90. Swearing
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

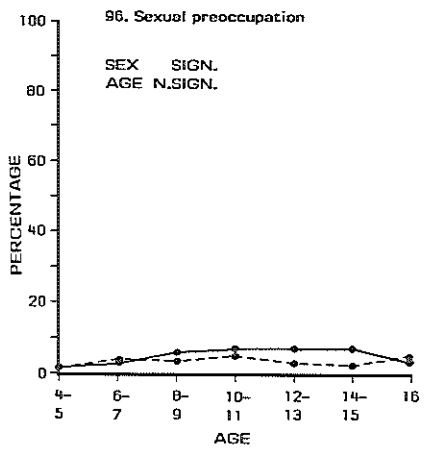

91. Suicidal talk
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

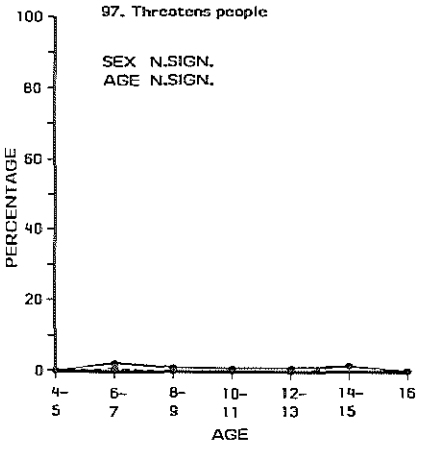

92. Talks or walks in sleep
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

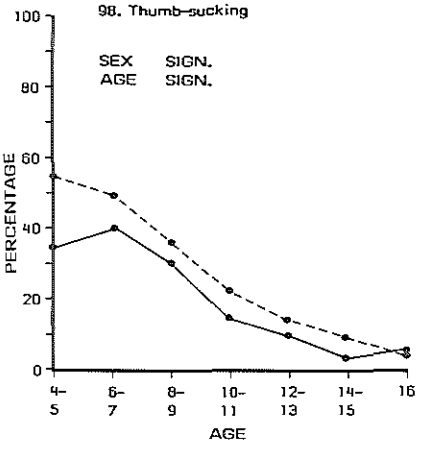

93. Talks too much
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

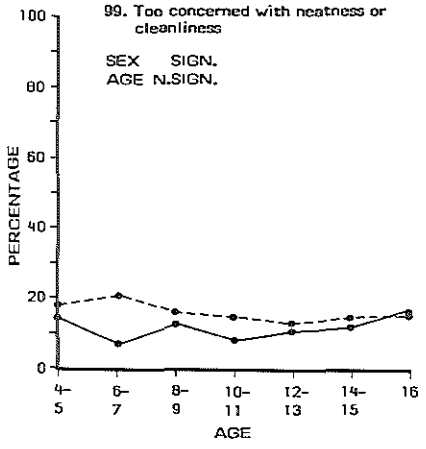

94. Teases a lot
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

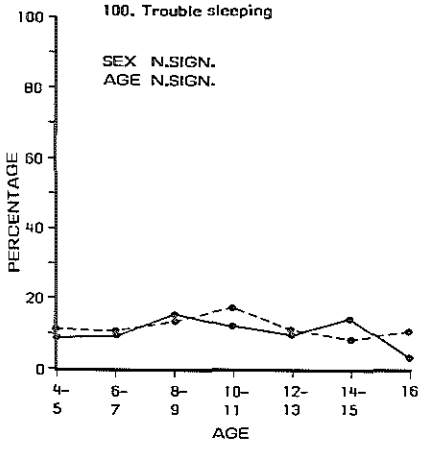

95. Temper tantrums
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

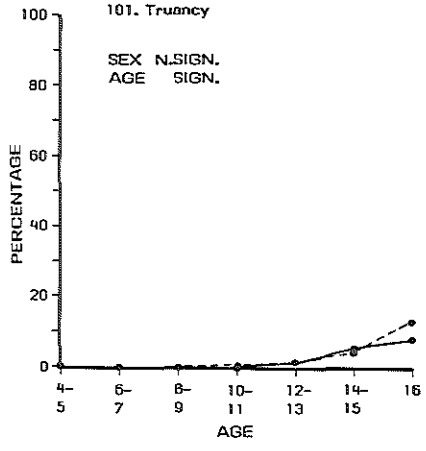

96. Sexual preoccupation
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

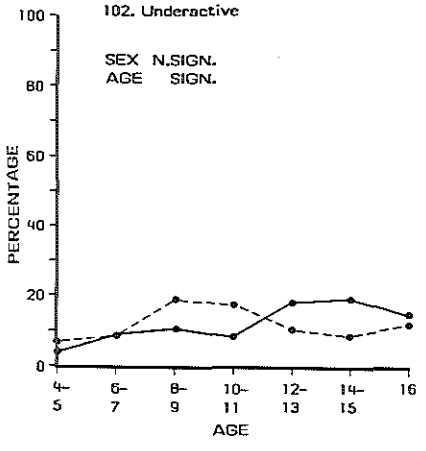

97. Threatens people
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

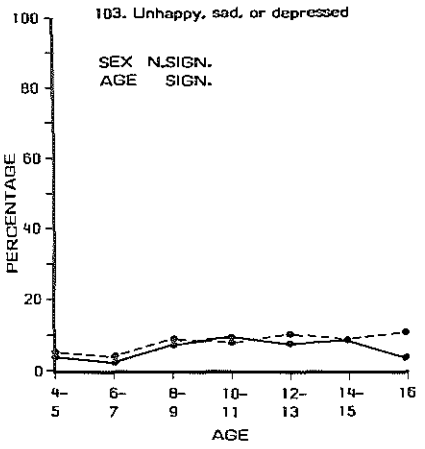

98. Thumb-sucking
SEX SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

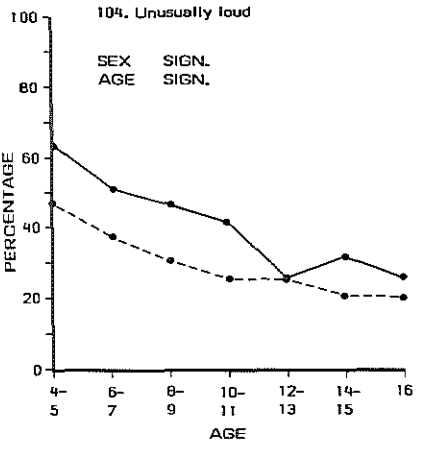

99. Too concerned with neatness or cleanliness
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

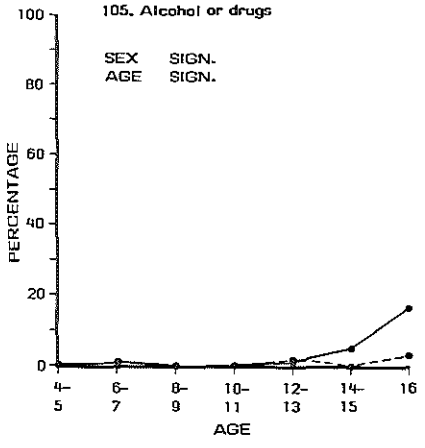

100. Trouble sleeping
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

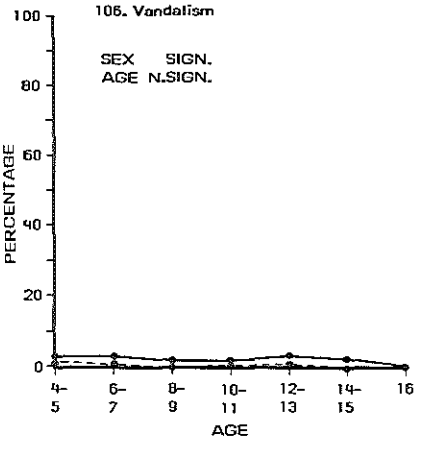

101. Truancy
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

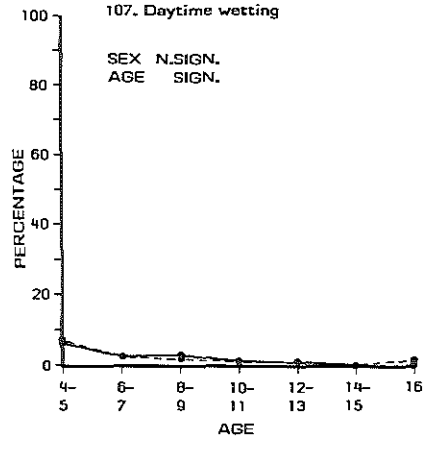

102. Underactive
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

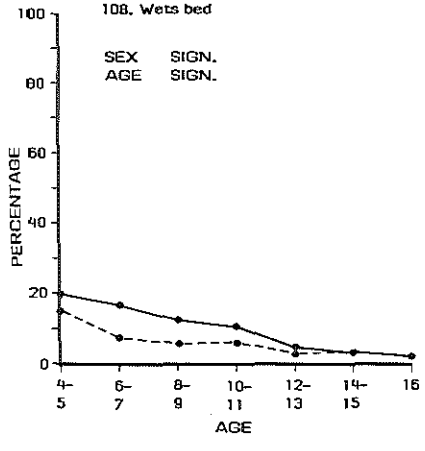

103. Unhappy, sad, or depressed
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
AGE

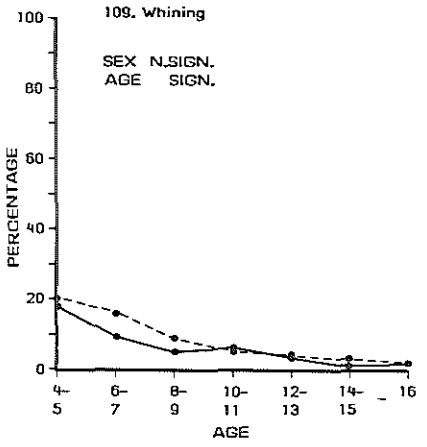

104. Unusually loud
SEX SIGN.
AGE SIGN.
PERCENTAGE
AGE

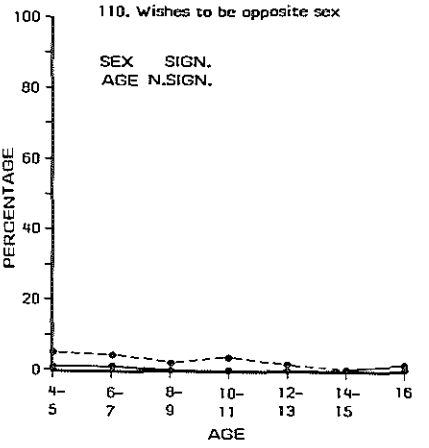

105. Alcohol or drugs
SEX SIGN.
AGE SIGN.
PERCENTAGE
AGE

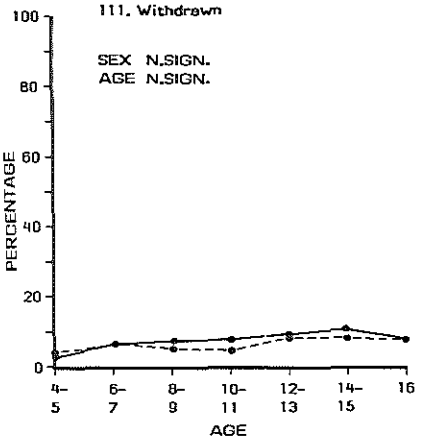

106. Vandalism
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
AGE

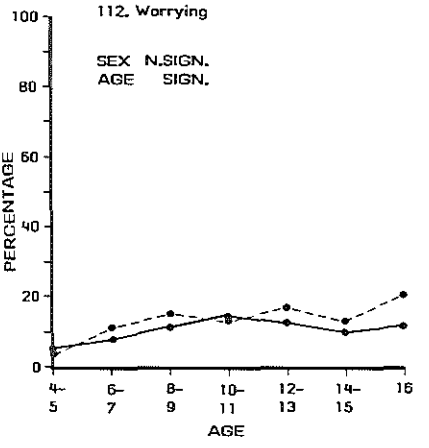

107. Daytime wetting
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
AGE


108. Wets bed
SEX SIGN.
AGE SIGN.
PERCENTAGE
AGE


109. Whining
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE


110. Wishes to be opposite sex
SEX SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE


111. Withdrawn
SEX N.SIGN.
AGE N.SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE


112. Worrying
SEX N.SIGN.
AGE SIGN.
PERCENTAGE
100
80
60
40
20
0
4-5
6-7
8-9
10-11
12-13
14-15
16
AGE

APPENDIX C

Mean total behavior problem score by age and sex for the general population sample (N=2076) and the nonreferred sample (N=2033)

General population sample

Age group (yrs)	Both sexes Mean	Both sexes Standard deviation	Boys Mean	Boys Standard deviation	Girls Mean	Girls Standard deviation
4	23.5	13.3	24.4	13.8	22.5	12.7
5	21.9	14.2	23.1	13.6	20.8	14.7
6	21.1	15.1	23.2	17.1	19.0	12.6
7	22.7	15.9	25.4	16.4	20.3	15.1
8	23.1	17.4	23.8	18.4	22.4	16.4
9	24.3	18.7	26.4	20.6	22.2	16.5
10	24.0	18.0	24.7	18.7	23.4	17.5
11	18.6	14.7	20.1	16.6	17.2	12.6
12	19.7	17.2	21.5	18.4	17.9	15.9
13	20.1	16.4	20.5	15.9	19.7	16.9
14	17.1	14.5	18.1	15.0	16.2	14.1
15	18.3	16.5	19.9	17.3	16.6	15.6
16	16.0	14.8	15.4	13.5	16.7	15.9

Nonreferred sample

Age* (yrs)	Both sexes Mean	Both sexes Standard deviation	Boys Mean	Boys Standard deviation	Girls Mean	Girls Standard deviation
4	23.4	13.5	24.0	14.2	22.8	12.8
5	21.3	14.0	22.5	13.2	20.2	14.6
6	21.4	15.1	23.0	17.4	20.1	12.7
7	22.3	16.0	25.5	16.1	19.5	15.5
8	21.5	16.7	21.5	17.0	21.5	16.5
9	24.0	17.0	25.6	18.9	22.5	14.9
10	22.3	16.6	24.6	17.7	20.0	15.1
11	19.7	16.8	19.6	17.5	19.7	16.2
12	19.2	16.8	20.4	17.5	18.1	16.0
13	18.5	14.4	20.0	15.4	17.0	13.1
14	17.1	14.5	16.4	12.8	17.7	15.7
15	18.6	15.9	20.6	16.0	16.6	15.8
16	15.3	13.6	15.2	13.5	15.5	13.9

*For this normative(nonreferred) sample actual age was used.

Percentage of variance accounted for by significant ($P < 0.01$) effects of clinical status,[a] gender and age[b] in ANCOVAs of behavior problems

Item		Clin.[c]	Gender	Age
1.	Acts too young	10	$<1^{M}$	2
2.	Allergy	$<1^{\star}$		$1^{NL\star}$
3.	Argues a lot	9		2^{Y}
4.	Asthma	$<1^{\star}$		
5.	Behaves like opposite sex		1^{F}	
6.	Encopresis	4	$1^{M\star}$	2^{Y}
7.	Bragging	1	12^{M}	2^{Y}
8.	Can't concentrate	14	1^{M}	
9.	Obsessions	17		$1^{NL\star}$
10.	Hyperactive	6	2^{M}	2^{Y}
11.	Too dependent	14		2^{Y}
12.	Lonely	12		
13.	Confused	14		
14.	Cries a lot	9		3^{Y}
15.	Cruel to animals	3	$<1^{M}$	
16.	Cruel to others	11	$<1^{M}$	
17.	Day-dreams	9		
18.	Harms self	2		
19.	Demands attention	17		4^{Y}
20.	Destroys own things	6	1^{M}	4^{Y}
21.	Destroys others' things	7	1^{M}	3^{Y}
22.	Disobedient at home	13	$<1^{M}$	4^{Y}
23.	Disobedient at school	10	$<1^{M}$	
24.	Doesn't eat well	3	$<1^{F}$	3^{Y}
25.	Poor peer relations	24		
26.	Lacks guilt	15		
27.	Easily jealous	8		4^{Y}
28.	Eats nonfood	<1		
29.	Fears	6	$<1^{F}$	6^{Y}
30.	Fears school	6		
31.	Fears impulses	3		
32.	Needs to be perfect	$<1^{\star}$		3^{NL}
33.	Feels unloved	17		
34.	Feels persecuted	16		3^{NL}
35.	Feels worthless	12		2^{0}
36.	Accident prone	2	1^{M}	1^{Y}
37.	Fighting	3	5^{M}	3^{NL}
38.	Is teased	14	$<1^{M\star}$	3^{NL}
39.	Hangs around with other children who get in trouble	4		2^{0}
40.	Hears things that aren't there	3		
41.	Impulsive	13	$<1^{M}$	
42.	Likes to be alone	<1		4^{0}
43.	Lying or cheating	10	$<1^{M\star}$	
44.	Bites fingernails	<1		2^{0}
45.	Nervous	21		3^{NL}
46.	Nervous movements	13		
47.	Nightmares	5		2^{Y}
48.	Not liked	14		$<1^{NL\star}$
49.	Constipated	1		
50.	Too fearful or anxious	16		
51.	Dizzy	3		3^{0}
52.	Feels too guilty	6		
53.	Overeating	<1		2^{0}
54.	Overtired	7		
55.	Overweight	$<1^{\star}$	$<1^{F}$	
56.[a]	Aches or pains	6		
56.[b]	Headaches	4		4^{0}
56.[c]	Nausea, feels sick	6		
56.[d]	Eye problems	2		
56.[e]	Skin problems	<1		
56.[f]	Stomach aches, cramps	4	1^{F}	
56.[g]	Vomiting	1		

Item		Clin.	Gender	Age
57.	Attacks people	7	$<1^{M}$	1^{NL}
58.	Picking	3		3^{Y}
59.	Plays with sexparts in public	1	$<1^{M}$	3^{Y}
60.	Plays with sexparts too much	2	$<1^{M}$	
61.	Poor schoolwork	15		4^{O}
62.	Clumsy	14	1^{M}	
63.	Prefers older children	3		3^{Y}
64.	Prefers younger children	10	1^{F}	
65.	Refuses to talk	12		
66.	Compulsions	5		
67.	Runs away from home	4		
68.	Screams a lot	9		4^{Y}
69.	Secretive	20		
70.	Sees things that aren't there	5		3^{Y}
71.	Self-conscious	3		
72.	Sets fires	2		1^{NL}
73.	Sexual problems	3		
74.	Showing off	7	2^{M}	6^{Y}
75.	Shy or timid	3	$<1^{F}$	
76.	Sleeps little	6		2^{NL}
77.	Sleeps much	<1		1^{NL}
78.	Smears feces	1		
79.	Speech problem	4	1^{M}	2^{Y}
80.	Stares blankly	12		
81.	Steals at home	7		
82.	Steals outside home	4		
83.	Hoarding	<1		5^{Y}
84.	Strange behavior	14		
85.	Strange ideas	8		
86.	Stubborn	12		
87.	Moody	14		
88.	Sulks a lot	7	$<1^{F\star}$	2^{Y}
89.	Suspicious	11		
90.	Swearing	4	2^{M}	
91.	Suicidal talk	4		
92.	Talks/walks in sleep	<1		
93.	Talks too much	2		3^{Y}
94.	Teases a lot	6	3^{M}	
95.	Temper tantrums	10	2^{M}	
96.	Sexual preoccupation	2	$<1^{M\star}$	
97.	Threatens people	3		
98.	Thumbsucking		$<1^{F}$	12^{Y}
99.	Too neat	$<1^{\star}$	$<1^{F}$	
100.	Trouble sleeping	10		
101.	Truancy	4		$6^{O\star}$
102.	Underactive	8		1^{O}
103.	Depressed	22		1^{O}
104.	Unusually loud	6	2^{M}	5^{Y}
105.	Alcohol or drugs		$<1^{M}$	6^{O}
106.	Vandalism	5	$<1^{M}$	
107.	Daytime wetting	3		3^{Y}
108.	Wets bed	3	$<1^{M}$	4^{Y}
109.	Whining	7		5^{Y}
110.	Wishes to be opposite sex		$<1^{F}$	
111.	Withdrawn	13		
112.	Worrying	15		2^{O}
Total behavior problem score		40	$<1^{M}$	3^{Y}

Note: Numbers in table indicate percentage of variance in scores of each item accounted for by each independent variable where the effect was significant at $P < 0.01$.

a For clinical status effects, ANCOVAs were performed on the sample containing normal (N=2033) and referred (N=1387) children
b For gender- and age effects ANCOVAs were performed on the general population sample (N=2076)
c Clinically referred children obtained higher scores on all behavior problems
F =higher scores for females M =higher scores for males
O =higher scores for older children Y =higher scores for younger children
NL=age effect not linear
★Indicates for each of the significant main effects those 5 effects having the smallest F values (see text)

APPENDIX E

Mean total social competence scores by age and sex for the general population sample (N=2076) and the nonreferred sample (N=2033)

General population sample

Age group (yrs)	Both sexes Mean	Both sexes Standard deviation	Boys Mean	Boys Standard deviation	Girls Mean	Girls Standard deviation
4	13.1	4.0	12.6	4.2	13.7	3.7
5	14.2	3.8	13.0	4.0	15.3	3.4
6	20.6	3.9	19.6	3.8	21.6	3.8
7	20.7	3.9	19.9	4.0	21.4	3.6
8	21.1	3.8	20.8	4.0	21.4	3.5
9	21.1	3.8	21.1	3.9	21.0	3.8
10	21.0	3.8	20.5	4.1	21.4	3.4
11	21.3	3.5	20.4	3.6	22.1	3.3
12	21.1	4.3	20.9	4.3	21.3	4.4
13	20.7	4.1	20.6	4.2	20.9	4.0
14	19.9	4.4	19.3	4.5	20.3	4.3
15	20.1	4.2	19.6	4.7	20.5	3.5
16	20.2	4.5	20.2	4.4	20.2	4.6

Nonreferred sample

Age* (yrs)	Both sexes Mean	Both sexes Standard deviation	Boys Mean	Boys Standard deviation	Girls Mean	Girls Standard deviation
4	13.0	3.8	12.5	4.1	13.4	3.5
5	14.3	3.8	13.1	4.0	15.3	3.5
6	20.7	3.6	20.1	3.7	21.2	3.5
7	20.9	4.0	19.7	3.9	21.8	3.8
8	21.1	3.8	20.9	4.1	21.3	3.4
9	21.1	3.9	20.8	4.0	21.3	3.8
10	21.1	3.9	20.7	4.1	21.5	3.6
11	21.5	3.4	20.7	3.5	22.1	3.3
12	21.0	4.0	21.1	3.9	20.9	4.1
13	20.8	3.8	20.5	3.8	21.0	3.8
14	20.1	4.4	19.5	4.6	20.6	4.3
15	20.0	4.4	19.6	5.1	20.3	3.6
16	20.2	4.4	20.4	4.3	20.0	4.5

*For this normative (nonreferred) sample actual age was used.

APPENDIX F

Percentage of variance accounted for by significant ($P < 0.01$) effects of clinical status[a], gender and age[b] in ANCOVAs of Social Competence

Item		Clin.	Gender	Age
I	A Number of sports	2		8^{NL}
	B Participation in sports	4		3^{NL}
	C Skill in sport	3		3^{NL}
II	A Number of nonsport activities		3^{F}	6^{Y}
	B Participation in activities			
	C Skill in activities	$<1^{\star}$	$<1^{F\star}$	
III	A Number of organizations	$<1^{\star}$	$<1^{F\star}$	11^{O}
	B Participation in organizations	3		9^{O}
IV	A Number of jobs	6^{R}		8^{O}
	B Job performance	2^{R}		12^{O}
V	1 Number of friends	8		
	2 Contacts with friends	14		$2^{Y\star}$
VI	A Behavior with siblings	6		$2^{NL\star}$
	B Behavior with other children	14		2^{NL}
	C Behavior with parents	9		3^{O}
	D Plays, works by self	1		
VII	1 Academic performance‡	12	1^{F}	2^{Y}
	2 (No) Special class‡	7	$<1^{F}$	
	3 (No) Grade repetition‡	1	$<1^{F}$	12^{Y}
	4 (No) Other school problems‡	28	2^{F}	
Total activities scale			$<1^{F}$	3^{O}
Total social scale		15	$<1^{F}$	4^{O}
Total school scale‡		22	2^{F}	5^{Y}
Total social competence score		12	$<1^{F}$	

a For clinical status effects, ANCOVAs were performed on the sample containing normal (N=2033) and referred (N=1387) children.

b For gender- and age effects, ANCOVAs were performed on the general population sample (N=2076).

R =higher scores for referred children. All other clinical status effects showed lower scores for referred children.

‡ =excludes 4- and 5-year-olds

F =higher scores for females O =higher scores for older children

M =higher scores for males Y =higher scores for younger children

*Indicates for each of the significant main effects those 2 having the smallest F values

APPENDIX G

Sensitivity and specificity for total behavior problem scores at various cutoff levels for each gender and age group

	Boys 4 - 5 yrs		Girls 4 - 5 yrs	
Total behavior problem score	Sensitivity,% Referred (N=124)	Specificity,% Nonreferred (N=153)	Sensitivity,% Referred (N=48)	Specificity,% Nonreferred (N=166)
17	91.9	38.6	93.7	43.4
18	90.3	41.2	89.6	45.2
19	88.7	47.1	85.4	47.6
20	87.9	50.3	83.3	51.8
21	86.3	52.3	79.2	54.2
22	85.5	54.2	79.2	59.6
23	84.7	56.9	72.9	63.9
24	83.1	58.2	70.8	65.7
25	83.1	61.4	68.7	69.3
26	80.6	63.4	64.6	71.7
27	79.4	68.0	64.6	74.1
28	78.2	69.9	64.6	75.9
29	75.8	72.5	64.6	77.1
30	75.0	73.9	64.6	78.9
31	74.2	76.5	64.6	82.5
32	73.4	78.4	64.6	83.7
33	72.6	79.7	64.6	84.9
34	71.0	81.0	60.4	85.5
35	70.2	83.0	56.2	87.3
36	68.5	86.9	54.2	87.3
37	66.1	87.6	54.2	88.0
38	65.3	88.2	54.2	89.8
39	64.5	88.9	52.1	90.4
40	60.5	90.2	52.1	91.6
41	59.7	91.5	52.1	91.6

Total behavior problem score	Boys 6 - 11 yrs Sensitivity,% Referred (N=505)	Boys 6 - 11 yrs Specificity,% Nonreferred (N=454)	Girls 6 - 11 yrs Sensitivity,% Referred (N=245)	Girls 6 - 11 yrs Specificity,% Nonreferred (N=486)
19	93.5	49.8	90.2	57.8
20	92.5	53.1	89.4	60.5
21	91.5	54.4	88.2	62.6
22	90.3	57.3	86.5	65.8
23	89.7	60.6	85.7	67.9
24	88.9	63.0	84.5	70.2
25	87.7	65.0	83.7	71.6
26	86.7	66.1	83.3	73.5
27	85.3	67.4	80.0	74.7
28	83.6	70.3	78.8	75.5
29	83.4	72.2	76.7	76.1
30	82.8	74.0	75.9	78.0
31	82.0	75.3	74.3	78.4
32	80.6	77.3	73.5	79.8
33	79.2	78.0	71.8	82.3
34	77.6	80.6	71.4	83.3
35	75.8	81.3	69.8	85.2
36	74.9	81.7	68.6	86.4
37	72.9	82.2	67.3	87.7
38	71.7	83.5	63.7	88.3
39	70.3	85.2	61.2	88.7
40	68.9	86.3	59.6	89.7
41	68.1	87.9	56.3	90.3
42	67.3	89.2	54.7	90.5
43	66.1	89.6	52.7	90.5
44	64.6	90.1	51.8	91.2
45	63.2	90.5	50.2	92.0
46	61.6	91.0	49.4	92.8

Total behavior problem score	Boys 12 - 16 yrs Sensitivity,% Referred (N=288)	Boys 12 - 16 yrs Specificity,% Nonreferred (N=370)	Girls 12 - 16 yrs Sensitivity,% Referred (N=175)	Girls 12 - 16 yrs Specificity, % Nonreferred (N=388)
16	95.8	55.4	90.3	60.8
17	95.8	58.6	89.1	63.1
18	94.4	60.8	88.6	66.0
19	94.4	63.2	88.6	67.5
20	93.4	66.2	88.0	69.6
21	92.4	68.6	86.0	72.2
22	91.3	71.1	86.3	74.2
23	90.6	72.7	86.3	74.7
24	89.6	73.2	85.7	76.5
25	87.8	75.4	84.6	78.4
26	87.2	77.6	82.9	79.9
27	86.8	78.6	80.6	82.0
28	86.1	80.0	79.4	83.0
29	84.7	81.1	78.3	84.8
30	84.4	81.6	76.6	85.8
31	83.3	82.2	76.6	86.9
32	81.9	82.7	76.0	87.9
33	80.6	83.8	75.4	87.9
34	78.8	86.2	74.3	88.9
35	77.4	87.0	73.1	89.7
36	77.1	88.4	70.3	90.7
37	74.7	89.7	69.1	92.0
38	73.6	91.1	66.9	92.0
39	72.2	91.4	65.7	92.5
40	71.2	91.9	64.0	92.8

APPENDIX H

Sensitivity and specificity for total social competence scores at various cutoff levels for each gender- and age group

	Boys 4 - 5 yrs		Girls 4 - 5 yrs	
Total social competence score	Sensitivity, % Referred (N=113)	Specificity,% Nonreferred (N=141)	Sensitivity,% Referred (N=42)	Specificity,% Nonreferred (N=162)
7.0	37	93	31	96
7.5	42	88	31	96
8.0	45	83	33	95
8.5	47	82	36	94
9.0	50	80	40	93
9.5	53	77	50	90
10.0	55	73	52	89
10.5	58	70	55	88
11.0	61	66	57	83
11.5	65	65	60	80
12.0	69	60	67	78
12.5	72	54	67	76
13.0	76	49	69	71
13.5	77	44	74	66
14.0	79	40	76	57
14.5	81	36	81	49
15.0	85	30	83	43
15.5	86	26	83	35
16.0	88	21	88	30
16.5	89	16	90	26
17.0	91	15	90	23

Total social competence score	Boys 6 - 11 yrs Sensitivity,% Referred (N=406)	Specificity,% Nonreferred (N=385)	Girls 6 - 11 yrs Sensitivity,% Referred (N=199)	Specificity, Nonreferred (N=424)
13.0	23	97	13	98
13.5	27	96	17	98
14.0	29	95	18	97
14.5	33	94	20	96
15.0	37	91	22	95
15.5	40	89	23	93
16.0	43	87	26	92
16.5	46	84	28	91
17.0	48	81	32	90
17.5	51	78	34	89
18.0	54	75	41	85
18.5	57	70	45	83
19.0	59	66	49	80
19.5	65	62	51	76
20.0	69	58	56	70
20.5	73	51	60	64
21.0	77	47	63	59
21.5	79	41	66	54
22.0	82	37	70	49
22.5	84	32	72	41
23.0	85	26	77	34
23.5	87	21	81	29
24.0	89	17	86	23
24.5	92	15	88	20
25.0	95	11	90	16

Total social competence score	Boys 12 - 16 yrs Sensitivity,% Referred (N=263)	Specificity,% Nonreferred (N=348)	Girls 12 - 16 yrs Sensitivity,% Referred (N=162)	Specificity,% Nonreferred (N=363)
13.0	29	94	23	96
13.5	32	93	26	95
14.0	37	91	27	94
14.5	41	90	31	93
15.0	42	88	34	92
15.5	46	85	37	91
16.0	51	83	41	89
16.5	52	82	44	86
17.0	56	78	48	83
17.5	60	75	52	78
18.0	65	71	55	73
18.5	68	66	59	69
19.0	70	61	63	67
19.5	73	58	65	62
20.0	76	53	67	58
20.5	80	48	69	52
21.0	82	43	73	47
21.5	86	39	77	41
22.0	87	36	81	36
22.5	90	31	83	27
23.0	90	28	84	23
23.5	91	24	85	20
24.0	94	20	88	18
24.5	95	17	90	16

APPENDIX I

Dear parents or guardians,

So far a population survey on children's interests, their skills and behavior has never been carried out in the Netherlands. As staff members of the Sophia Children's Hospital, part of the University Hospital Rotterdam, we ask you to join in this population survey.
We want to get to know what children like to do in their leisure hours, their sports and hobbies, but we also want to collect data on a number of important parts of the children's everyday life such as sleeping, eating, use of medicine, bodily complaint, etc.
We want to collect these data from a great number of children in Zuid-Holland in the age of 4 to 16. Your child has been selected for this survey in an absolutely accidental way. What we ask of you is whether you are willing to receive a member of the staff of our university; he will ask you a number of questions concerning the above-mentioned child. The interview will take about half an hour and this staff member will call on you in the period from March to May of this year.
We hope that you want to cooperate because the results of this survey may contribute to the improvement of help to children who are in need of it.

Yours sincerely,

APPENDIX II

Dear Teacher,

You will find a questionnaire enclosed with questions about school records, social skills and the behavior of children.
As a staff member of Sophia Children's Hospital, which forms part of the Erasmus University Hospital, we ask your cooperation in a population survey which consists of filling in this questionnaire.
The aim of the survey is to collect data about the most important fields of the functioning of nursery school and primary school children in Zuid-Holland, who were selected via a so-called a-select test.
These data are of great significance in order to get an impression of the skills of children and also of the nature of potential problems. Besides data from the parents, we interviewed a short time ago, the information the teacher can provide is a source of information as to the functioning of children in the widest sense. Therefore we consider these data indispensable.
What we ask you is whether you are willing to fill in the questionnaire concerning the above-mentioned child. You will find enclosed a letter of consent of the child's parents or guardians and a return-envelope for returning the questionnaire.
We hope that you will send back the filled-in list because the results of this survey are to contribute to the improvement of aid to children who are in need of it.
Thanking you very much in advance.

Yours sincerely,

APPENDIX III

Dear parents/guardians,

You will remember thas some time ago you were interviewed. A great number of questions were asked you then about the behavior and the skill of your child. We wanted to get to know what sort of small and big problems parents with children have. Thanks to your cooperation, which has been of great value, we have got an impresssion of this.
However, we want to extend the survey even further. Therefore we come again to you with a request, namely whether you (mother, father or both) are willing to come to the Sophia Children's Hospital with your child. One member of the survey team will put a number of questions to you then, while one of the physicians will talk to your child about a great number of matters which are of importance in the child's daily life, such as school, friends, health, etc. There will also be some skill-tests. This survey is not at all burdening for the child. Most children rather like to talk about themselves and about what they do.
With the results of this survey we are able to inform people who are in daily contact with children such as general practitioners, pediatricians, teachers, etc. of the things which are important for children and how they can be helped, if necessary.
Of course we will refund the expenses made and besides there will be a small present for your child as a reward for your coming. You will find enclosed a letter for your child explaining the aim of the survey. You may give it to your son or daughter.
If you are willing to come, which is of great importance for this survey, we are prepared to accomodate ourselves to the time that suits you and your child best. Before long one of us will phone to ask how you feel about it.

With kind regards,

APPENDIX IV

Dear

We want to ask you the same question we asked your parents namely whether you want to cooperate in a survey. It is about this: all boys and girls have to deal with things they can do well, but sometimes they also meet with small or big problems. On television or radio there are sometimes programs in which boys and girls talk about themselves or about common problems, but then it is mostly about one subject. It seems important to us to find out how a big group of boys and girls of different ages get on with matters concerning school, home, friends, hobbies, sports, etc. We ask this in order to get a good impression of the things young people have or don't have difficulties with. May be you wonder: Why do they find this so important? We find it so important because in this way we can imform doctors, teachers and others who are in daily contact with young people, of what boys and girls feel to be important and of the things you can help them best with, if necessary. You can help to bring this about.
What we ask you is whether you, together with your mother or father, or the three of you, are willing to come to us. One of us will have a talk with your mother and/or father and I or an other doctor will talk to you and you will be asked to do so some small tests. You will be asked questions about yourself and you may also tell something if you feel like it.
We hope that you will join in.
Good-bye,

Yours sincerely,

REFERENCES

Achenbach TM. The classification of children's psychiatric symptoms: a factor-analytic study. Psychol Monogr 1966, 80 (7), 1-37.

Achenbach TM. The child behavior profile: I. Boys aged 6-11. J Consult Clin Psychol 1978, (a), 46, 478-488.

Achenbach TM. Psychopathology of childhood: research problems and issues. J Consult Clin Psychol, 1978, (b), 46, 759-776.

Achenbach TM. The child behavior profile: an empirically based system for assessing children's behavioral problems and competencies. Int J Ment Health 1979 (a), 7, 24-42.

Achenbach TM. What is child psychiatric epidemiology the epidemiology of? In: Earls F ed. Studies of children. New York: Prodist 1980 (a), 96-117.

Achenbach TM. DSM-III in light of empirical research on the classification of child psychopathology. J Am Acad Child Psychiatry 1980 (b), 19, 395-412.

Achenbach TM, Edelbrock CS. The classification of child psychopathology: a review and analysis of empirical efforts. Psychol Bull 1978, 85, 1275-1301.

Achenbach TM, Edelbrock CS. The child behavior profile: II. Boys aged 12-16 and girls aged 6-11 and 12-16. J Consult Clin Psychol 1979, 47, 223-233.

Achenbach TM, Edelbrock CS. Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. Monogr Soc Res Child Dev 1981, 188.

Achenbach TM, Edelbrock CS. Manual for the child behavior checklist and revised child behavior profile. Queen City: Queen City Printer Inc, 1983.

American psychiatric association. Diagnostic and statistical manual of mental disorders. 3rd ed. Washington DC, APA, 1980.

Bartko JJ. On various intraclass correlation reliability coefficients. Psychol Bull 1976, 83, 762-765.

Bille B. Migraine in school children: a study of the incidence and short-term prognosis, and a clinical, psychological and encephalographic comparison between children with migraine and matched controls. Acta Paediatr 1962, 51, Suppl 136.

Cassel JM. Potentials and limitations of epidemiology. In: Katz AH, Felton JS, eds. Health and the community: readings in the philosophy and sciences of public health. New York: The Free Press, 1965, 432-445.

Cederblad M. A child psychiatric study on Sudanese Arab children. Acta Psychiatr Scand 1968, 44, Suppl 200.

Centraal bureau voor de statistiek (CBS). Bevolking der gemeenten van Nederland op 1 januari 1981. 's Gravenhage, 1981.

Cohen, J. A coefficient of agreement for nominal scales. Educ Psychol Measurement 1960, 20, 37-46.

Cohen J. Statistical power for the behavioral sciences. Rev Ed. New York, Academic Press, 1977.

Cohen-Matthijsen TA. Slaapstoornissen bij kinderen. Lisse, Swets & Zeitlinger, 1981.

Connell HM, Irvine L, Rodney J. The prevalence of psychiatric disorders in rural school children. Aust NZ J Psychiatry 1982, 16, 43-46.

Conners CK. Rating scales for use in drug studies with children. Psychopharmacol Bull 1973, 9 (special issue), 24-84.

Cox A, Rutter M, Yule B, Quinton D. Bias resulting from missing information: some epidemiological findings. Br J Prev Soc Med 1977, 31, 131-136.

Cullen KJ, Boundy CAP. Factors relating to behaviour disorders in children. Aust Peadiatr J 1966 (a), 2, 70-80.
Cullen KJ, Boundy CAP. The prevalence of behaviour disorders in the children of 1,000 Western Australian families. Med J Austr 1966 (b), 53, 805-808.
Das P, Doorn H. Volumebeleid Wijkwelzijnsplanning, C.B.S.-buurten Rotterdam naar achterstand. Bijlage 1a. Rotterdam: Gemeentelijk Bureau voor Onderzoek en Statistiek, 1980.
Davie R, Butler H, Goldstein, H. From birth to seven: the second report of the National Child Development Study (1958 Cohort). London; Longman, 1972.
Earls F. Prevalence of behavior problems in 3-year-old children; a cross-national replication. Arch Gen Psychiatry 1980, 37, 1153-1157.
Edelbrock CS, Achenbach TM. The teacher version of the child behavior profile. I. Boys aged 6-11. J Consult Clin Psychol 1984, 52, 207-217.
Eme RF. Sex differences in childhood psychopathology: review. Psychol Bull 1979, 86, 574-595.
Epidemiogisch preventief onderzoek Zoetermeer (EPOZ). Vijfde voortgangsverslag. Deel 1. Algemeen. Rotterdam, Erasmus Universiteit, Medische Faculteit, Instituut Epidemiologie, 1979.
Feild HS, Armenakis AA. On use of multiple tests of significance in psychological research. Psychol Rep 1974, 35, 427-431.
Feshbach S. Aggression. In: Mussen PH, ed. Carmichael's Manual of child psychology. Vol. II, 3rd ed. New York, Wiley, 1970, 159-259.
Fogelman K. Britain's sixteen-year-olds: preliminary findings from the third follow-up of the National Child Development Study (1958 Cohort). London: National Children's Bureau, 1976.
Glidewell JC, Mensh IN, Gildea MCL. Behavior symptoms in children and degree of sickness. Am J. Psychiatry 1957, 114, 47-53.
Goldberg D, Huxley P. Mental illness in the community: the pathway to psychiatric care. London: Tavistock, 1980.
Gould Schwartz M, Wunsch-Hitzig R, Dohrenwend B. Estimating the prevalence of childhood psychopathology; a critical review. J Am Acad Child psychiatry 1981, 20, 462-476.
Graham, P, Rutter M. The reliability and validity of psychiatric assessment of the child: II. Interview with the parent. Br J Psychiatr 1968, 114, 581-592.
Graham PJ. Epidemiologic perspectives on maladaptation in children: neurological, familial, and social factors. J Am Acad Child Psychiatry 1978, 17, 197-208.
Graham P, Stevenson J. A twin study of genetic influences on behavioral deviance. J Am Acad Child Psychiatry 1985, 24, 33-41.
Gross RT, Dornbusch SM. Disordered processes of elimination; Enuresis. In: Levine MD, Carey WB, Crocker AC, Gross RT, eds. Developmental - behavioral pediatrics. Philadelphia: Saunders 1983, 573-586.
Hodges K, Kline J, Stern L, Cytryn L, McKnew D. The development of a child assessment interview for research and clinical use. J Abnorm Child Psychol 1982 (a), 10, 173-189.
Hodges K, McKnew D, Cytryn L, Stern L, Kline J. The child assessment schedule (CAS) diagnostic interview: a report on reliability and validity. J Am Acad Child Psychiatry 1982 (b), 21, 468-473.
Hughes JG. Synopsis of pediatrics. 6th ed. Saint Louis: Mosby, 1984.
de Jonge GA. Kinderen met enuresis: een epidemiologisch en klinisch onderzoek. Proefschrift Rijksuniversiteit Utrecht. 1969. Assen: Van

Gorcum, 1969.
Kaffman M, Elizur E. Infants who become enuretics: a longitudinal study of 161 Kibbutz childern. Monogr Soc Res Child Dev 1977, 170, 14-40.
Kastrup M. Psychic disorders among preschool children in a geographically delimited area of Arhus County Denmark. Acta Psychiatr Scan 1976, 54, 29-42.
Kastrup M. Urban - rural differences in 6-year olds. In: Graham PJ, ed. Epidemiological approach in child psychiatry. London; Academic Press, 1977, 181-195.
Kellam SG, Branch JD, Agrawal KC, Ensminger ME. Mental health and going to school: the Woodlawn program of assessment, early intervention, and evaluation. Chicago: University of Chicago Press, 1975.
Pringle MLK, Butler NR, Davie R. 11,000 seven-year-olds: first report of the National Child Development Study (1958 Cohort). London: Longmans, 1966.
Kendell RE. The role of diagnosis in psychiatry. Oxford: Blackwell, 1975.
Klein DF, Gittelman-Klein R. Problems in the diagnosis of minimal brain dysfunction and the hyperkinetic syndrome. Int J Ment Health 1975, 4, 45-60.
Krupinski J, Baikie AG, Stoller A, Graves J. O'Day DM, Polke P. A community health survey of Heyfield, Victoria. Med J Austr 1967, 54, 1204-1211.
Langner TS, Gersten JC, McCarty Ed, et al. A screening inventory for assessing psychiatric impairment in children 6 to 18. J Consult Clin Psychol 1976, 44, 286-296.
Lapouse R, Monk MA. An epidemiologic study of behavior characteristics in children. Am J Public Health 1958, 48, 1134-1144.
Lapouse R, Monk MA. Fears and worries in a representative sample of children. Am J Orthopsychiatry 1959, 29, 803-818.
Lapouse R, Monk MA. Behavior deviations in a representative sample of children; variation by sex, age, race, social class and family size. Am J Orthopsychiatry 1964, 31, 436-446.
Lapouse R. The relationship of behavior to adjustment in a representative sample of children. Am J Public Health 1965, 55, 1130-1141.
Leslie SA. Psychiatric disorders in the young adolescent of an industrial town. Br J Psychiatry 1974, 125, 113-124.
Lilienfeld AM. Foundations of epidemiology. New York, Oxford University Press, 1976.
Lindeman RH, Merenda PF, Gold RZ. Introduction to bivariate and multivariate analysis. Glenview: Scott, Foresman, 1980, 193-195.
Links PS, Stockewell M, Abichandani F, Simeon J. Minor physical anomalies in childhood autism. Part I. Their relationship to pre- and perinatal complications. J Autism Dev Disord 1980, 10, 273-285.
Maccoby EE, Jacklin CN. Sex diffences in aggression: a rejoinder and reprise. Child Dev 1980, 51, 964-980.
McCoy SA. Clinical judgement of normal childhood behavior. J Consult Clin Psychol 1976, 44, 710-714.
McGee R, Silva PA, Williams S. Behaviour problems in a population of seven-year-old children; prevalence, stability and types of disorder: a research report. J Child Psychol Psychiatry 1984, 25, 251-259.

van der Meeren WBJM, Verbrugge HP. Youth health care in The Netherlands 0-19 years. Leidschendam: Ministry of Health and Environmental Protection, 1981.
Miller FJW, Court SDM, Knox EG, Brandon S. The school years in Newcastle upon Tyne 1952-62: being a further contribution to the study of a thousend families. London: Oxford University Press, 1974.
Miller LC. Q-sort agreement among observers of children. Am J Orthopsychiatry 1964, 34, 71-75.
Mitchell S, Shepherd M. A comparative study of children's behaviour at home and at school. Br J Educ Psychol 1966, 36, 248-254.
Nunnally JC. Psychometric theory. New York: McGraw-Hill, 1967.
O'Leary KD, Bennett Johnson S. Psychological assessment. In: Quay HC, Werry JS, eds. Psychopathological disorders of childhood. 2nd ed. New York: Wiley, 1979, 210-247.
Orvaschel H, Sholomskas D, Weissman M. Assessing children in psychiatric epidemiologic studies. In: Earls F, ed. Studies of children. New York: Prodist, 1980; 84-96.
Quay HC. Classification. In: Quay HC, Werry JS, eds. Psychopathological disorders of childhood. 2nd ed. New York; Wiley 1979. 1-43.
Quay HC, Peterson DR. Interim manual for the revised behavior problem checklist. In: Applied social sciences. Coral Gables: University of Miami. 1983.
Quinton D, Rutter M, Rowlands O. An Evaluation of an interview assessment of marriage. Psychol Med 1976, 6, 577-586.
Remschmidt H, Hohner G, Merschmann W. Epidemiology of delinquent behaviour in children. In: Graham PJ, ed. Epidemiological approaches in child psychiatry. London: Academic Press, 1977, 253-274.
Richman N, Graham PJ. A behavioural screening questionnaire for use with three-year-old children; preliminary findings. J Child Psychol Psychiatry 1971; 12, 3-33.
Richman N, Stevenson JE, Graham PJ. Prevalence of behaviour problems in 3-year-old children: an epidemiological study in a London borough. J Child Psychol Psychiatry 1975, 16, 277-287.
Richman N, Stevenson J, Graham PJ. Pre-school to school: a behavioural study. London: Academic Press, 1982.
Robins LN. Follow-up studies. In: Quay HC, Werry JS, eds. Psychopathological disorders of childhood. 2nd ed. New York: Wiley, 1979, 483-515.
Routh DK, Schroeder CS, O'Tuama LA. Development of activity level in children. Dev Pathol 1974, 10, 163-168.
Rutter M, Graham P. Psychiatric disorder in 10- and 11-year-old children. Proc R Soc Med 1966, 59, 382-387.
Rutter M. A children's behaviour questionnaire for completion by teachers: preliminary findings. J Child Psychol Psychiatry 1967, 8, 1-11.
Rutter M, Graham P. The reliability and validity of the psychiatric assessment of the child: I. Interview with the child. Br J Psychiatry 1968, 114, 563-579.
Rutter M, Yule B, Quinton D, Rowlands O, Yule W, Berger M. Attainment and adjustment in two geographical areas: III. Some factors accounting for area differences. Br J Psychiatry 1974, 125, 520-533.
Rutter M, Cox A, Tupling C, Berger M, Yule W. Attainment and adjustment in two geographical areas: I. The prevalence of psychiatric disorders. Br J Psychiatry 1975, 126, 493-509.

Rutter M, Graham P, ChadwicK OFD, Yule W. Adolescent turmoil: fact or fiction? J Child Psychol Psychiatry 1976, 17, 35-56.
Rutter M. Surveys to answer questions: some methodological considerations. In: Graham PJ, ed. Epidemiological approaches in child psychiatry. New York: Academic Press, 1977, 1-31.
Rutter M. Invulnerability, or why some children are not damaged by stress. In: Shamsie SJ, ed. New directions in children's mental health. New York: SP Medical & Scientific Books, 1979, 53-75.
Rutter M, Tizard J, Whitmore K. Education, health and behaviour. Huntington: Krieber, 1981 (first published in 1970)
Schaie KW. A general model for the study of developmental problems. Psychol Bull 1965, 64, 92-107.
Shaywitz SE. The Yale neuropsychoeducational assessment scales. Schizophr Bull 1982, 8, 360-424.
Shepherd M, Oppenheim AN, Mitchell S. Childhood behaviour disorders and the child guidance clinic; an epidemiological study. J Child Psychol Psychiatry 1966 (a), 7, 39-52.
Shepherd M, Oppenheim AN, Mitchell S. The definition and outcome of deviant behaviour in childhood. Proc R Soc Med 1966 (b), 59, 379-382.
Shepherd M, Oppenheim B, Mitchell S. Childhood behaviour and mental health. London: University of London Press, 1971.
Silverstein AB. Validity of WISC-R short forms. J Clin Psychol 1972, 31, 696-697.
Steinhausen HC, Rentz A, Gobel D. Kindliche Verhaltensauffaligkeiten im Eltern- und Lehrerurteil. Acta Paedopsychiatr 1983, 49, 61-71.
Tinsley HEA, Weiss DJ. Interrater reliability and agreement of subjective judgment. J Consult Psychol 1975, 22, 358-376.
Weiss G, Hechtman L. The hyperactive child syndrome. Science 1979, 205, 1348-1354.
Werner EE, Bierman JM, French FE. The children of Kauai: a longitudinal study from the prenatal period to age ten. Honolulu: University of Hawaii Press, 1971.
Werry JS. Studies on the hyperactive child: IV. An empirical analysis of the minimal brain dysfunction syndrome. Arch Gen Psychiatry 1968, 19, 9-16.
Werry JS, Quay HC. The prevalence of behavior symptoms in younger elementary school children. Am J Orthopsychiatry 1971, 41, 136-143.
van Westerlaak, JM, Kropman JA, Collaris JWM. Beroepenklapper. 2 delen. Nijmegen: Instituut voor Toegepaste Sociologie, 1975.

SUMMARY IN THE DUTCH LANGUAGE

In hoofdstuk I wordt de geschiedenis van de kinderpsychiatrische epidemiologie beschreven. Het onderzoek is verricht (1) om prevalentie gegevens te verkrijgen over gedrags- en emotionele problemen en over vaardigheden van kinderen in de algemene bevolking, (2) om verschillen vast te stellen die verband houden met demografische variabelen, (3) om gegevens te verkrijgen met betrekking tot de validiteit van de gedragsvragenlijst voor kinderen 4-16 jaar (Child Behavior Checklist, CBCL) - een vragenlijst die ontworpen is om gestandaardiseerde gegevens te verzamelen over een groot aantal klinisch relevante gedragingen die kinderen kunnen vertonen, (4) om Nederlandse gegevens te vergelijken met gegevens van andere onderzoeken die in de literatuur vermeld worden, vooral met de gegevens van het onderzoek van Achenbach e.a. (1981), (5) om gegevens die verkregen zijn met behulp van de gedragsvragenlijst te vergelijken met andere onderzoekmethoden, met name het kinderpsychiatrisch onderzoek.

In hoofdstuk II wordt een literatuuroverzicht gegeven, vooral gericht op de concepten en methoden zoals deze gehanteerd worden in de kinderpsychiatrische epidemiologie. De grootste obstakels worden gevormd door de diversiteit en willekeurigheid waarmee het begrip psychiatrische stoornis gedefinieerd wordt, en het tekort aan instrumenten die getoetst zijn in populaties die verschillen van die waarvoor het instrument oorspronkelijk ontworpen was.

In hoofdstuk III worden de methoden van de eerste stap van het onderzoek besproken. Er werd van een representatieve steekproef van 2076 Zuid-Hollandse kinderen van 4 tot en met 16 jaar door de ouders een CBCL ingevuld door middel van een thuis-interview. De response was 85,1%. Een "normale", niet-aangemelde groep werd verkregen door verwijdering uit het bestand van 43 kinderen die in het jaar voorafgaande aan het interview waren verwezen naar een instelling voor geestelijke volksgezondheid. Bovendien werd voor 73% van de 4- tot en met 11-jarige kinderen van de startpopulatie informatie verkregen door middel van de TRF (Teacher Report Form), een instrument analoog aan de CBCL. Verder werden door 1387 ouders van kinderen die aangemeld werden bij 21 instellingen voor geestelijke gezondheidszorg een CBCL ingevuld bij de intake. Intraclass correlatiecoefficienten werden berekend voor test-hertest en tussenbeoordelaar betrouwbaarheid, evenals voor de betrouwbaarheid wanneer twee ouders de lijst onafhankelijk van elkaar invulden. De ICC's varieerden van 0,69 tot 0,99.

In hoofdstuk IV worden de resultaten van de eerste stap vermeld. De frequenties voor ieder CBCL-item afzonderlijk zijn grafisch weergegeven. ANCOVA's verricht teneinde de effecten en interacties van leeftijd, geslacht, en sociaal economische klasse (SEK) in de algemene bevolkingssteekproef te berekenen voor ieder gedragsprobleem en sociale competentie item van de CBCL, lieten een groot aantal verschillen zien. Gedragsproblemen neigen af te nemen met toenemende leeftijd. Voor jongens werden slechts iets hogere gemiddelde gedragsprobleem totaalscores gevonden dan voor meisjes. Veel meer uitgesproken was de tendens bij jongens hoger

te scoren op sociaal ongewenst gedrag. Hoe lager de SEK, hoe groter het aantal gedragsproblemen en hoe lager de sociale competenties. Ouders uit lagere sociaal economische klassen gaven hun kinderen hogere scores op gedragsprobleem items behorend tot het Externalizing syndroom. Het percentage kinderen in de niet-verwezen groep (N=2033) die ooit op school waren blijven zitten nam sterk toe na de overgang van de lagere school naar het vervolgonderwijs. Het onderscheidend vermogen waarmee de CBCL verwezen en niet-verwezen kinderen kan onderscheiden, uitgedrukt in het percentage misclassificatie, bedroeg 18,9%. Er werd gebruik gemaakt van de cutoff scores corresponderend met de 90e percentiel van de gedragsprobleem scores en de 10e percentiel van de sociale compententie scores van de niet-verwezen groep. Weging van items op basis van hun discriminant functies verbeterde het discriminatief vermogen van de CBCL niet. Voor de meeste items die het sterkst geassocieerd waren met het verwezen zijn naar een instelling voor geestelijke gezondheid, gold dat dit ook het geval was in het onderzoek van Achenbach e.a.

In hoofdstuk V worden de bevindingen besproken. Hollandse ouders neigden er meer toe hun kinderen hoger te scoren op items die geassocieerd zijn met "moeilijk gedrag", dan ouders in buitenlandse onderzoeken. De gemiddelde totaalscores van de gedragsprobleem-items en sociale competentie-items totaalscores zoals vermeld in Achenbach's onderzoek kwamen overeen met die in dit onderzoek. De gegevens die in dit onderzoek vermeld worden zijn van belang voor de Nederlandse situatie, aangezien de CBCL het eerste instrument is dat ontworpen is om een grote verscheidenheid aan gedragsproblemen bij kinderen te beschrijven en dat genormeerd is voor de nederlandse bevolking. De relevantie van een groot aantal leeftijds, geslachts en SEK effecten voor andere onderzoeken en theorieen, werd besproken.

In hoofdstuk VI worden de methoden van de tweede, meer klinisch georienteerde, stap besproken. In dit gedeelte van het onderzoek, werden 8- en 11-jarigen geselecteerd voor klinische evaluatie indien hun totaalscores op de CBCL en/of de TRF op of boven de 85e percentiel van hun respectievelijke cumulatieve frequenties uitkwamen. Uit de overige 85% van ieder van de twee leeftijdsgroepen werd aselect een "normale" groep getrokken en eveneens klinisch geevalueerd. Van de 153 geselecteerde kinderen werden er 116 (75,8%) tesamen met de ouder(s) onderzocht. Het onderzoek van het kind bestond uit een gestruktureerd kinderpsychiatrisch onderzoek, de WISC-R short form, coordinatie testjes, en het maken van 2 tekeningen. Gegevens van de ouders over het kind, evenals gegevens over een aantal persoonlijke- en gezinskenmerken, werden door middel van een semi-gestruktureerd interview verkregen. Onmiddellijk nadat de psychiater het kind uitgebreid onderzocht had, werd door hem de ernst en de aard van de psychiatrische problematiek genoteerd. De uiteindelijke scoring van de ernst van de aard van de problematiek gebaseerd op alle schriftelijke informatie van de onderzoeker die het kind had gezien en van de interviewers die de gesprekken met de ouders hadden gevoerd, werd uitgevoerd door drie kinderpsychiaters. Tussen-beoordelaar en tussen-interviewer betrouwbaarheden waren goed.

In hoofdstuk VII worden de resultaten gegeven. De Kendall correlatie coefficient tussen de CBCL en het oordeel op basis van het gestruktureerde kinderpsychiatrisch onderzoek is 0,42, terwijl de correlatie tussen de TRF en het kind-onderzoek 0,28 is. De correlatie tussen TRF en CBCL is 0,26. De prevalenties van kinderpsychiatrische problemen in de algemene bevolking van 8- en 11-jarigen werden geschat. Zeven procent van de 8- en 11-jarigen werden als ernstig gestoord beoordeeld, terwijl geschat werd dat 26% matige of ernstige problematiek vertoonde. De onderzochte groepen waren te klein om prevalenties van specifieke stoornissen te geven. ANOVA's en discriminant functie analyse werden uitgevoerd ten einde die CBCL items te verkrijgen die het beste onderscheid aangaven tussen kinderen die klinisch-psychiatrisch als probleemkinderen werden beschouwd en kinderen uit de vergelijkingsgroep die geen noemenswaardige psychiatrische problemen vertoonden. Veel items die in de tweede stap van het onderzoek als goed onderscheidende items naar voren kwamen, bevonden zich ook tussen de items, die in de eerste stap een goed onderscheid maakten tussen verwezen en niet-verwezen kinderen. Een aantal sociale- en gezinsfactoren bleken geassocieerd te zijn met psychiatrische problematiek van het kind.

In hoofdstuk VIII worden de resultaten van de tweede stap besproken. Er wordt aandacht besteed aan de conclusies die verbonden kunnen worden aan de lage correlaties tussen onderzoekers die kinderen in verschillende situaties beoordelen. Het gewicht dat toegekend moet worden aan beoordelingen vanuit verschillende informatiebronnen verschilt met het soort probleem dat het kind vertoont. Er wordt geconcludeerd dat er meer onderzoek nodig is om voor verschillende soorten stoornissen of problemen de relatieve waarde te bepalen van verschillende soorten gegevens. De tamelijk hoge prevalentie van kinderpsychiatrische problemen die gevonden werd in dit onderzoek vergeleken met andere onderzoeken is deels toe te schrijven aan de arbitraire aard van het klinisch oordeel. De gezins- en sociale factoren die in dit onderzoek geassocieerd bleken met psychiatrische problematiek ondersteunen bevindingen van ander onderzoek.

In hoofdstuk IX wordt het gehele onderzoek geevalueerd en worden er aanbevelingen gedaan voor verder onderzoek. Vooral wordt er aandacht besteed aan de praktische gevolgen die onze gegevens hebben met betrekking tot het gebruik van de CBCL.

ACKNOWLEDGEMENTS IN THE DUTCH LANGUAGE

Dit onderzoek was nooit tot stand gekomen zonder het werk en de steun van velen. Het is mij dan ook een genoegen om diegenen die mij behulpzaam waren bij het tot stand komen van dit proefschrift te danken.

Mijn promotor en opleider Prof.Dr.J.A.R.Sanders-Woudstra ben ik veel dank verschuldigd voor het enthousiasme, de zorg en de daadwerkelijke steun waarmee zij dit proefschrift begeleidde. Haar veelzijdige belangstelling en grote klinische kennis inspireerden niet alleen dit proefschrift maar vormden ook de basis van mijn klinische en psychotherapeutische vorming in de kinderpsychiatrie.

Prof.Dr.H.A.Valkenburg dank ik voor de vele waardevolle en vooral praktische adviezen die hij als tweede promotor vanaf een vroege fase van het onderzoek gaf. Ik denk met genoegen terug aan de vele uren besteed aan het doornemen van het manuscript waarin mij duidelijk werd dat "vrij associeren" met researchgegevens ook nuttig is.

I am very grateful to have come into contact with Prof.T.M.Achenbach who generously gave his time to teach me the principles of empirical research in the scientific field of developmental psychopathology. He was the inspiror of this study. Through his knowledge and his talented teaching but most of all through his enthusiasm and integrity as a researcher he gave me the firm belief that it is worth while to work hard in order to obtain empirical data in this still largely unreclaimed and fascinating research area.

Prof.Dr.J.W.Schudel ben ik dank verschuldigd voor de kritische opmerkingen die hij als lid van de promotiecommissie heeft willen maken.

Wetenschappelijk onderzoek zal voor mij altijd verbonden blijven met degene die mij met de praktijk ervan in contact bracht, namelijk Prof.Dr.A.van Harreveld. Aan hem heb ik te danken dat ik op het "California Institute of Technology" de sfeer heb mogen proeven hoe research in zijn pure vorm wordt bedreven.

Verder dank ik allen die aan mijn klinische vorming maar ook aan de persoonlijk kant van mijn opleiding hebben bijgedragen of dit nog steeds doen.

Grard Akkerhuis heeft als psycholoog-methodoloog vanaf het begin van het onderzoek meegewerkt. Zonder de inzet waarmee ook hij zich in het avontuur van dit onderzoek stortte en zonder de deskundigheid die hij zich verwierf in het werken met grote bestanden was het onderzoek niet zo ver gekomen.

Francien Timmer heeft in de beginfase als methodologe aan het onderzoek meegewerkt.

Mijn collega Guy Berden dank ik voor zijn inzet bij het interviewen en het scoren van de gegevens. Niet alleen zijn deskundigheid maar vooral zijn toewijding waren een grote steun bij het onderzoek. Marijke Berden dank ik voor het zo belangeloos uitwerken van de scores op de intelligentie tests.

Monika Althaus werd als methodologe in een laat stadium met het onderzoek geconfronteerd. Desondanks hebben haar doortastende methodologische visie en haar heldere uitleg van de vaak moeilijke statistische materie een grote invloed uitgeoefend.

Jaantje Prince heeft zich gedurende alle fasen van het onderzoek als student-assistente ingezet en wist menigmaal weer orde in de administratieve chaos te brengen.

Annegret de Bel heeft met een uitstekend klinisch inzicht vele ouderinterviews afgenomen. Met genoegen denk ik terug aan de tijd dat wij samen de provincie introkken om de "thuis-interviews" af te nemen.

Hanneke van der Lee dank ik voor het vele speurwerk dat zij verrichtte om informatie over de non-responders te verkrijgen.

Ook Betty Glas en Hester Loyson-Kalf dank ik voor de vele ouderinterviews die zij op knappe wijze hebben verricht.

In de beginfase van het onderzoek heeft Karinke van der Leij-Kouwenhoven hard meegewerkt (zelfs in haar vakantie) aan de basis van het onderzoek, namelijk het trekken van de steekproef van Zuid-Hollandse kinderen.

Het benaderen van ouders van 2600 kinderen en het interviewen van 2076 van deze ouders is zeker geen kleinigheid geweest. Ik ben de volgende studenten dan ook zeer dankbaar voor het uitvoeren van deze soms zware taak: Ida Dielemans, Irene Donkers, Diane Jesurun, Barbro Kvaal, Hanneke van der Lee, Saskia van der Merwe, Elisabeth Pont die als medisch studenten meewerkten in het kader van hun keuzepraktikum, en Walther van Gogh, Fernando Miguel, Mieke Mijnlieff, Hein Tweehuysen, Jan van Veen en Anita Verburg die als psychologie studenten in het kader van hun onderzoekstage meewerkten.

Annemarie Illsley-de Jonge dank ik voor de uiteindelijke vorm die dit proefschrift kreeg. Kenmerkend voor de zorgvuldigheid waarmee zij het typewerk verrichte was het feit dat zij het manuscript in de brandkluis liet opbergen ! Echter, tekstverwerkers storen zich niet aan brandkluizen, met als gevolg dat de ondankbare taak wachtte een groot deel van het manuscript dat verloren was gegaan over te typen.

Ook de anderen van het secretariaat dank ik voor het vele werk dat in de loop van het onderzoek verricht is: Marianne Breugem voor het tikken van de vele adressen en het uitgeven van de vragenlijsten, Geke Kuiper-van Dellen voor haar bereidheid zich bezig te houden met wat de puntjes op de "i" leken te zijn, en Inge Putter-Demmendaal voor het typen van vele vragenlijsten en haar steun in de eindfase.

Drs.H.J.A.Schouten van de afdeling Biostatistiek dank ik voor zijn adviezen over de betrouwbaarheidsmaten.

Henriette van Steensel, destijds werkzaam op de afdeling epidemiologie, dank ik voor de praktische adviezen bij het trekken van de steekproef.

Dank gaat ook uit naar Dr.A.Hofman en Dr.J.P.Vandenbroucke van de afdeling epidemiologie voor hun waardevolle adviezen.

John de Kuyper en Hillie Versprille van de audiovisuele dienst dank ik voor het vele werk aan de figuren en tabellen.

Ellen Nelemans-van de Broek dank ik voor de vriendelijke en nauwgezette wijze waarmee zij de literatuurlijst opstelde.

Uiteraard wil ik de ouders van de kinderen en ook de kinderen zelf die aan het onderzoek meewerkten heel hartelijk danken voor hun enthousiaste medewerking. De openhartigheid en vriendelijkheid waarmee zij ons een blik gunden in hun leefwereld heeft een diepe indruk gemaakt.

Mijn collega's op de afdeling kinderpsychiatrie waren mij tot steun door hun bereidwilligheid mij te ontlasten van een aantal taken.

Maarten Bezemer en Jan Bruijn dank ik niet alleen voor het doornemen van het manuscript, maar vooral voor het feit dat er tussen het begin van de medische studie en dit laatste academische diploma vele dierbare herinneringen liggen.

Mijn broer, George Verhulst, dank ik voor de nauwlettendheid waarmee hij de tekst heeft gezuiverd van taalkundige fouten. De gulheid waarmee hij zijn vrije tijd belangeloos opofferde waardeer ik bijzonder.

Mijn ouders hebben mij de ruimte gegund en gegeven om mij te kunnen ontplooien. Hun stimulering en steun bij het opdoen van nieuwe ervaringen oefenen ook nu nog steeds hun invloed uit.

Tenslotte bedank ik mijn vrouw Mia, die op bewonderenswaardige wijze het klimaat wist te scheppen waarin dit stuk werk verricht kon worden. Ook onze kinderen Francien, Susanne en Paul, ging het schrijven van dit proefschrift niet ongemerkt voorbij, hetgeen onder andere bleek uit de Sinterklaas surprise die bestond uit een proefschrift voor pappa met daarin een heuse wetenschappelijke verhandeling.

CURRICULUM VITAE

De schrijver van dit proefschrift werd op 18 augustus 1951 te Rotterdam geboren. Hij bezocht het Gemeentelijk Lyceum te Beverwijk en later het Rotterdamsch Lyceum, waar hij in 1968 het eindexamen HBS-B behaalde. Hij studeerde geneeskunde aan de Medische Faculteit Rotterdam en werd in 1975 tot arts bevorderd. Van februari tot augustus 1971 werkte hij op basis van een "special graduateship" op de afdeling neurofysiologie van het California Institute of Technology te Pasadena, USA (Prof.Dr.A.van Harreveld). Verder was hij in 1971 gedurende enkele maanden als student-assistent verbonden aan de afdeling Chemische Pathologie van de Medische Faculteit Rotterdam, waar hij in 1972 nog enige tijd op vrijwillige basis werkzaam was. Gedurende zijn militaire dienst was hij van 1975 tot 1976 werkzaam als keuringsarts. Zijn opleiding tot psychiater vond plaats in het Academisch Ziekenhuis te Rotterdam (Prof.Dr.G.A.Ladee; voor de stage kinderpsychiatrie Prof.Dr.J.A.R.Sanders-Woudstra; voor de stage neurologie Prof.Dr.A.Staal) van begin 1976 tot eind 1980. Sedert januari 1981 is hij als staflid verbonden aan de afdeling kinder- en jeugdpsychiatrie van het Academisch Ziekenhuis Rotterdam/Sophia Kinderziekenhuis (hoofd:Prof.Dr.J.A.R.Sanders-Woudstra). Sinds 1980 is hij opleidingskandidaat van de Nederlandse Vereniging voor Psychoanalyse.