CHAPTER 1.

GENERAL INTRODUCTION

In 1990, at the commencement of the project from which this thesis arises, it could be said that there was little or no data on the effectiveness of the mental health treatments being delivered to children and adolescents in Australian hospital or community health settings, or in fact any other Australian setting. At this time, the debate with regard to the development of child mental health services and policies needed to take place without the benefit of relevant data. It is obviously problematic when judgements about the likely cost-effectiveness of health initiatives are made in the absence of firm evidence of effectiveness but in the presence of detailed information about cost. Under these circumstances, lower-cost options can appear more attractive, even though outcomes might turn out to be less satisfactory in the long term. This is particularly problematic where children and adolescents are concerned - the social and economic costs of poor social, emotional and behavioural adjustment in young people compound as they become adults, often repeating into the next generation familial patterns of psychological dysfunction, or substance abuse, or patterns of emotional neglect and abuse which perpetuate a hostile, alienated attitude to the community.

During the 1990s - the decade during which the project from which this thesis arises was conducted - there emerged a strong increase in interest in the empirical basis of clinical treatments in medicine and the social sciences. In the field of psychotherapy and behaviour change, at least, this did not lead to a proportional increase in evaluations of what was taking place in clinical settings. Instead, the debate with regard to the development of child mental health services and policies came to be framed by the demands from those in academic settings that only treatment types already shown to be effective in their randomised controlled trials
(RCT) should take place in clinical settings. In this way, the whole basis of the debate shifted, with evidence of the effectiveness of various specific recommended treatments - mostly behavioural and cognitive-behavioural - being strongly put forward.

However, with regard to the clinical group at the centre of the present thesis – the most severely emotionally disturbed children and adolescents – very little evaluation of treatment programs catering to them had taken place at all. There was little change in this state of affairs to be seen in the 1990s. This was despite general agreement that seriously emotionally disturbed children and adolescents are an "at risk" group, and that ongoing evaluation and research into the effectiveness of services provided for them is important. Not only were mental health programs for this group introduced without a requirement for its effectiveness to be monitored, programs were closed without reference to objective measures of their effectiveness, or ineffectiveness. As in periods previous to the “empirical treatment” era of the 1990s, changes in treatment programs took place according to the social trends of the times. The social trends of the 1990s in Australia, as in the US and UK earlier, moved against longer-term, more intensive, day or residential, psycho-social treatment for young people; shorter-term, less intensive, outpatient, psychopharmacological interventions came into favour.

The demands on the institutional settings for the treatment of the most severely emotionally disturbed youths to demonstrate their effectiveness - if not cost-effectiveness - is well reflected in the titles of books of the 1990s. For example, Child Psychiatric Units at the Crossroads (Chesson and Chisholm, 1996) and The Traumatic Bond Between the Psychotherapist and Managed Care (Weisgerber, Bronfman, Loula, Mitchell & Wolf, 1999) reflect the struggle for survival of institutions and professional practices from earlier eras. Other titles promised “advances” in outcome research (Psychiatric treatment: Advances in outcome research, Mirin,
Gossett & Grob, 1991) yet failed to provide any new empirical results, developments in methodology, or conceptual advances. The titles (and summaries on the back cover and preface) may have emphasised “research” (e.g. *Inpatient Child Psychiatry: Modern practice, research and the future*, Green & Jacobs, 1998) yet the publications provided very superficial content in this respect. The “detailed, up-to-date information on their functions, treatment processes and outcomes” readily recognised to be necessary to support “arguments for units’ survival” (ix, from the preface, Chesson and Chisholm, 1996) failed to emerge.

As the 1990s drew to a close, it was concluded in a comprehensive book-length survey of the efficacy of child and adolescent treatments - *What Works for Whom: a Critical Review of Treatments for Children and Adolescents* by Fonagy, Target, Cottrell, Phillips & Kurtz (2002) - that “specialist residential or day care facilities” – that is, the facilities where the most intensive treatments for the most severely emotionally disturbed children have traditionally taken place – show “a startling and worrying lack of evidence to guide practice” (p. 390).

A reason often given for this lack of evidence is that most child and family mental health settings lack the resources and knowledge to develop and implement an ongoing system for auditing clinical outcomes which is reliable, valid and free of observer bias while remaining operationally simple and relatively inexpensive. It is an irony of the “empirical treatment” era that the traditional specialist facilities have become all but extinct in the Anglosphere without the collection and examination of evidence with regard to its effectiveness or ineffectiveness.

At Arndell Child and Adolescent Unit, a combined health and educational facility for severely emotionally disturbed children jointly administered by the Department of Health and the Department of Education in NSW, Australia, a program of evaluative research was implemented
in 1990 which continued until 2001. During this time clinical data was obtained on standardised instruments from all children and adolescents admitted to the inpatient and daypatient programs at registration, during the course of treatment, and at 6 months and 12 months after discharge. The evaluative research program was not designed to compare the effectiveness of the traditional, psychodynamic milieu-based treatment, dominant in the setting until the mid-90s, against the contemporary, “empirically-validated”, cognitive behavioural treatments which superseded it. Yet, the program forms the basis of an interesting “natural experiment” not anticipated when it was conceived by the candidate.

This thesis has three main aims.

The first aim (Chapter 2) is to outline the main areas of conceptual, pragmatic and methodological confusion and neglect which impede progress in research in this area. This chapter challenges the conventional wisdom of the 1990s that the randomised controlled trial (RCT) is sine qua non of scientific method. It argues for the plurality of data analytic strategies and research designs, and therefore the legitimacy of the attempt to develop methods such as those employed in the present study.

The second aim (Chapter 3) is to critically review the reported findings about the effectiveness of treatments of this general type (intensive day and residential treatment in specialist facilities), and reported findings about the predictors of good outcomes for this treatment type. This chapter confirms the conclusion of Fonagy, Target, Cottrell, Phillips & Kurtz (2002), cited above, that there is very little to guide practice. It argues for the necessity of studies which address some basic questions, and therefore the necessity of studies of the present type.
Having argued for the legitimacy of its methods and the necessity to address basic questions, the third aim of this thesis (Chapters 4, 5 & 6) is to report the results of a study drawn from the accumulated data of the evaluative research program at Arndell Child and Adolescent Unit. The study addresses the question: *what child, family and treatment variables predict outcome for the children and adolescents treated at Arndell Child and Adolescent Unit from 1990 to 1999?* The results of this study challenge the conventional wisdom of the times in which it took place. It shows, first of all, that a relatively modest evaluative research program of this type can provide substantive results of theoretical and clinical interest. It also shows that the traditional psychodynamic milieu-based treatment taking place in this institution was more effective than the contemporary “empirically-validated” cognitive-behavioural treatment which replaced it.
CHAPTER 2.
LITERATURE REVIEW:
CONCEPTUAL, PRAGMATIC AND METHODOLOGICAL ISSUES

2.1 Introduction to this Chapter

A critical examination of the empirical studies of intensive psycho-social interventions for severely emotionally and behaviourally disturbed children & adolescents reveals serious methodological weaknesses and logical inconsistencies. They undermine the veracity of reported substantive findings to such a degree that it is quite difficult to get a sense of what we confidently know about this area. This is so with regard even to basic issues such as are these interventions effective? or what child, family or treatment factors are associated with a good outcome? While it may seem harsh, this is not an idiosyncratic or controversial summary position to adopt, and anticipated in Chapter 1 with reference to the conclusions of Fonagy, Target, Cottrell, Phillips & Kurtz (2002).

The methodological problems evident in the studies published in this area have been discussed in numerous reviews. Studies of the effectiveness of day treatment programmes have been reviewed by, for example, Zimet & Farley (1985), Baenen, Parris-Stephens & Glenwick (1986), and Sayegh & Grizenko (1991), and studies of the effectiveness of residential treatment programmes have been reviewed, for example, by Durkin & Durkin (1975) and Pfeiffer & Strzelecki (1990). However, while identifying many of the more gross methodological shortcomings, their criticisms generally do not go far enough, and some of their conclusions (Zimet & Farley, 1985; Sayegh & Grizenko, 1991) are unjustifiably positive given the small number of well-conducted studies with appropriately analysed results demonstrating positive treatment effects. When the same confusions with regard to basic issues appear time and again
in reviews of the area, perpetuating misunderstandings which become codified and incorporated within a “standard practice”, progress remains impeded.

In this thesis, an attempt is made to move beyond a catalogue of research studies and their shortcomings against an “ideal” research methodology. The present chapter is concerned with identifying and outlining the main recurring conceptual and methodological confusions in the field of evaluative research. While this requires a much broader focus than the evaluation of specialist residential or day care facilities – the problems encountered are representative of the broader field of evaluative research – the illustrations will be drawn from child and adolescent studies where possible. The following chapter (Chapter 3) then allows a more compact review of the substantive findings in this area than would otherwise be possible.

This chapter is organised around four identified areas of confusion or neglect.

The confused notion that there are particular, “prescribed” research designs to which scientific inquiry needs to adhere [2.2].

The neglect of important pragmatic issues in conducting research in child units, such as staff resistance to self-evaluation [2.3].

Confusions around the notion of the “clinical” as opposed to the “statistical” significance of outcomes [2.4].
Confusions around the notion of “outcome” itself, and the problems that arise from the failure to maintain (a) distinctions between logically different concepts of outcome, and (b) distinctions between different sources of outcome. [2.5]
2.2 “Prescribed” research design

In their review of studies of the effectiveness of day treatment programmes, Sayegh & Grizenko (1991) baldly recommend the use of control groups. In their review of studies of the effectiveness of residential treatment programmes Pfeiffer & Strzelecki (1990) recommend "employing powerful statistical techniques and sophisticated designs" (p.852) in a similarly unqualified way.

The main problem with recommendations of this type lies in the implication that sound research methodology consists of the application of a set of standard procedures rather than the development of procedures which most adequately address the question of interest, the phenomena of interest, and the nature of the data generated or available. These recommendations fail to recognise that a control group does not in itself represent either sound or unsound research methodology; it is only in relation to a particular approach taken to a particular research question, and the logical consequences which follow for the interpretation of the results, that the value of a particular type of control group can be reckoned.

2.2.1 Matching research design to the research question

To illustrate, if the research question pertains to treatment effectiveness, and the model of data analysis is the conventional statistical test of difference between group means, an experimental or quasi-experimental (Cook & Campbell, 1979) research design is required¹. Either type of

¹ Given the way in which the term “quasi-experimentation” has sometimes been used to refer loosely to different types of variation from experimental design, it is necessary to clarify that quasi-experimental design (Cook & Campbell, 1979) involves the comparison of the treatment group with another group but without random allocation - the comparison (or “control”) group often being “naturally occurring”. An experimental design - or to use the term adopted in the medical literature, a random controlled trial (RCT) - involves the random allocation of subjects/patients to either the treatment group or at least one other alternative group for the purposes of comparison. The difference between these two designs lies in the much more rigorous internal validity afforded by the latter, on one hand, and the pragmatic
control group design is suited to the application of statistical tests of group-differences. On the other hand, the application of this model of data analysis in the absence of a control group does not provide meaningful results. Studies which, for example, report the $p$ value of a t-test applied to the difference between the pre-treatment and the post-treatment means in a group and assert that “statistically significant” change due to treatment has been demonstrated rightly come under criticism in the reviews mentioned. There is no disagreement with that here.

However, if the research question pertains to predictors of treatment outcome, the required model of data analysis and research design are quite different. A well-established approach is to examine the relationship among predictors and outcome in multiple linear regression analyses of data from a naturalistic study. That is, the analysis is performed on outcomes arising in an actual treatment setting without the changes to clinical procedures entailed by the formation of treatment groups by random allocation (i.e. an experimental design). Nor does a control group external to the clinical setting need to be located (i.e. a quasi-experimental design) for meaningful analyses to take place.

In fact, if the scientific aim is understanding the emotional and behavioural outcomes for young people from various clinical, socio-demographic and family backgrounds following various life events, a naturalistic design utilising measures which intrude as little as possible is a logical requirement. These types of studies include epidemiological studies in the general population, as well as studies of predictors of outcome for the smaller special populations undergoing treatments. A naturalistic design is a logical necessity for addressing research questions of the advantages the former often offers, on the other. With an experiment it is difficult to raise logical alternatives to the conclusion that any difference in outcomes between the two groups is due to the different treatment conditions. With a quasi-experimental design it becomes easier to study the effects of variables to which individuals cannot readily be randomly allocated (eg. exposure to carcinogens, parental psychopathology).
second type\(^2\), just as control group designs (experimental or quasi-experimental designs) are a logical necessity for addressing research questions of the first type (i.e. are the treatments effective).

2.22 Matching research design to the phenomena under study?

The previous illustration pertains to the importance of matching research design to the research question at hand, and confusions associated with the mistaken assumption of a general superiority of comparison designs over correlational designs irregardless of the question at hand. There are also confusions around the topic of matching research design to the phenomena under study.

For example, some among those opposing RCTs in evaluating the effectiveness of psychological treatments argue that controlled studies are inherently unsuitable for this purpose. The most prominent arguments fall into two main types. One, that the random allocation of patients to treaters (or treatments) necessarily precludes important aspects of the psychotherapeutic process. Two, that the adoption of a RCT restricts the study to clinically trivial phenomenon.

The first type of argument is a specific case of the second, more general, type of argument and Seligman (1995) is perhaps its most prominent proponent. As will be shown, there are merits in the specificity of his argument, in contrast to the broad-brushed confusions of others who engage in less well specified arguments. Seligman has championed the “Consumer Reports”

\(^2\) It might be argued that questions of the second type – what are the predictors of outcome – can be addressed in control group designs where, for example, groups are formed on the basis of different values of predictor variables. This, however, overlooks the difference between research aimed at identifying predictors of outcome (i.e. what are the predictors of outcome) – such as the study reported in the present thesis - and research aimed at examining more closely the relationship between an already identified predictor and outcome (i.e. how does the identified predictor variable affect outcome).
survey (1994, 1995) – an US equivalent to the Australian “Choice” magazine – involving retrospective self-reports by consumers of their experience of counselling and mental health services. He believes “the main methodological virtue” of the Consumer Reports survey to be “its realism: it assessed the effectiveness of psychotherapy as it is actually performed in the field with the population that actually seeks it, and it is the most extensive, carefully done study to do this” (p. 971).

Seligman seems prepared to entirely overlook the risk of taking self-reports at face value. (We will return to this topic below, with reference to the classic McCord, 1978, study). He states, provocatively, that “random assignment, the prettiest of the methodological niceties in efficacy studies, may turn out to be worse than useless for the investigation of the actual treatment of mental illness in the field” (p. 974). It is often difficult in the current US debate about RCTs cast in terms of “effectiveness” versus “efficacy” studies to understand what positions are being taken by proponents of either side with regard to research design, and this is particularly so with the more general type of argument against the RCT. In contrast, Seligman’s basis for his argument is made clear. He asserts that in the normal clinical situation the allocation of patient to therapist, or allocation of patient to treatment, is non-random\(^3\). He then asserts that this allocation is more likely to be “right” than allocation at random.

The first part of Seligman’s assertion will not be disputed. Although relatively neglected as a subject of formal study, it can be readily accepted that there are powerful economic and social determinants for the type of mental health service that an individual receives, and that the matching up of specific patients with specific therapists follows equally powerful personal

\(^3\) Although Seligman does not state it, the allocation of therapist to treatment modality is also non-random.
valencies which are clearly non-random. The second part of Seligman’s assertion cannot be accepted, however, as it is unjustifiably optimistic. It is not reasonable to assume that self allocations are more likely to lead to therapeutic success than random allocation. For example, an individual with a history of selecting inappropriate partners may well show a seemingly uncanny ability to select a series of inappropriate therapists and enact similarly unsatisfactory outcomes (eg. Shur, 1994).

Whether natural allocation results in better outcomes than random allocation of patient to therapist/treatment is of course an empirical question, which Seligman readily recognises.

“It is worth mulling over what the results of an efficacy or effectiveness study might be if half the patients with a particular disorder were randomly assigned and were compared with half the patients not randomly assigned” (p. 974).

While correctly suggesting that his hypothesis can be tested through a RCT, Seligman seems, however, to be unaware of his self-contradiction. He appears to be confusing a hitherto neglected factor in formal psychotherapy research which may yet be shown to be an important determinant of outcome - the process by which therapists and patients find each other - with an intrinsic limitation in the RCT design itself. He confuses pragmatic and logical issues, which is not uncommon in criticisms of experimental designs (c.f. Judd, Smith & Kidder, 1991, p. 95 - 98).

The second, more general, type of argument that RCTs are inherently unsuitable for evaluating the effectiveness of psychological treatment is that the adoption of a RCT restricts the study to
clinically trivial phenomenon. Stated conversely, it is argued that RCTs are not suited to the study of clinically significant phenomenon.

Silberschatz’s position in the Persons & Silberschatz (1998) debate, which appeared in the special section on “Empirically Supported Therapies” in the *Journal of Consulting and Clinical Psychology*, illustrates this stance.

“When a patient seeks therapy, the therapist must try to answer several basic questions: what is bothering the patient? What does the patient hope to accomplish in treatment? What has impeded the patient from achieving his or her goals? How can the therapist help this patient? I believe that certain kinds of research studies are capable of providing useful data to answer these fundamental clinical questions [examples provided]… However, RCTs do not provide any meaningful help in addressing these questions and, consequently, they have had very little impact on clinicians and on the practice of psychotherapy. Goldfried & Wolfe (1996) have similarly suggested that RCTs have had minimal impact on the practice of therapy because they simply do not address the issues that are most pertinent to the practicing therapist: how to treat patients who suffer from multiple disorders, how to treat underlying personality problems, how to resolve clinical impasses, and how or why do treatments work – that is, what are the actual mechanisms and processes of therapeutic change?” (p. 128)

As with Seligman’s position, this again confuses pragmatic and logical issues. Silberschatz raises many different questions not related to the question of the effectiveness of treatments, and suggests – probably correctly - that experimental designs are not appropriate for addressing
many of them. However, as Persons points out in her side of the debate, the types of questions asked by Silberschatz “do not address the same questions the RCTs address” (p. 133) - that is, whether a treatment is effective.

Silberschatz muddies the water with his general, not clearly spelled out association of RCTs with clinically trivial studies. If he was, for example, to point out the pragmatic difficulties in randomly allocating real patients to real treatments and how this may be associated with the trivial nature of the phenomenon studied in some RCTs, this would be meaningful. Similarly, if Silberschatz was to point out that the much fewer number of RCTs of psychodynamic therapies relative to cognitive-behavioural and behavioural therapies means not that the former is ineffective but that their effectiveness remains yet to be demonstrated, and then explore the extra difficulties in conducting RCTs of more sophisticated or complex treatments, this would also be meaningful. (Pragmatic issues are discussed in the next section of this chapter [2.3]). If he was to point out that a prescriptive attitude to clinical research where the design (RCT) is determined first and the types of phenomenon that would most readily fit this approach is sought later – the not infrequent case of the methodological tail wagging the substantive dog - there would be no argument with that here. Pointing out that the “single-case experiment”, generally recognised as a “logical equivalent” of RCTs (Chambless & Hollon, 1998, p. 14), is not at all orientation-neutral as claimed by Persons (p. 128) among others, would also be meaningful. (This point is discussed later in this chapter [2.42]).

But Silberschatz does not do any of these things. He appears, rather, to be engaging in a rather weak defence of the fact that there are fewer studies demonstrating the effectiveness of psychodynamic therapies with specific problem types relative to cognitive and behavioural therapies by provocatively suggesting that the question of treatment effectiveness is “clinically
useless” (p. 133). He appears not to acknowledge that the types of questions he values and endorses (i.e. the clinical processes which predict outcome) are of little clinical value unless the treatment is demonstrably effective. Persons points this out very clearly. To put it another way, the clinical value of a study of outcome predictors (i.e. an investigation of the factors which distinguish relatively poor outcomes from the relatively good outcomes within a treatment group) depend to some degree on where the range of outcomes lie relative to general or clinical population norms, and relative to the range of outcomes for alternative treatments. A study of factors which distinguishes relatively good from relatively poor outcomes in a treatment study which produces outcomes ranging from (at best) no change to (at worst) deterioration, is of less clinical interest than a study of factors which distinguishes relatively good from relatively poor outcomes in a treatment study which produces outcomes ranging from (at best) improvement to (at worst) no change. By clinical interest, I refer here to the immediate, pragmatic interest of developing effective treatments, distinct from the theoretical interest in identifying the active ingredients associated with producing deterioration - which may well lead later to important pragmatic applications (The issue of “clinical significance” of treatment outcomes is discussed later in this chapter [2.4]. The confusion between logically distinct meanings of the notion of outcome is also discussed later in this chapter [2.5]).

As Persons points out, in her side of the debate, there is nothing inherent in experimental design, or its essential feature of random allocation to treatment, to associate it with trivial variables or artificial settings. This is well illustrated by the classic Cambridge-Somerville Youth Study (McCord, 1978), the power of which arises from its experimental design. This was a study of the long-term effects of an intervention program commenced in 1939 for a mixed group of over 500 boys, some considered “normal” and some considered in danger of delinquency. Half were randomly assigned to a program of active case-management which included educational,
medical and psychiatric attention as well as participation in group activities, while the other half were assigned to a control group where the boys participated only in providing information about themselves. McCord successfully traced 488 of the 506 boys more than 30 years later, in middle age. The unexpected and discomforting main results of this study are, in brief, as follows. First, there was no difference on a wide variety of objective outcome measures between men who had participated in the program and those in the control group. Second, where there were statistically reliable differences, they favoured the non-treated group. Third, despite this, the majority of the treated group reported that treatment was helpful to them – a point which appears not to be considered by Seligman in respect of the Consumer Report study.

The random allocation of delinquent youth to either treatment or no-treatment groups means that there is no ready logical alternative to the conclusion that receiving treatment as youths led to poorer outcomes for these men in middle age.

Silberschatz’s general type of argument why a controlled experiment is not suitable to study the phenomena of psychotherapy effectiveness is based on a confusion between pragmatic and logical issues already seen in Seligman’s more specific type of argument. In fact, the broad style of argument in which Silberschatz engages suggests a form of resistance to inquiry, which is among the very significant pragmatic difficulties which need to be addressed and managed if evaluative research is to take place in clinical settings. These and other pragmatic issues are examined in the next section.
2.3 Pragmatic issues

Important pragmatic issues arise when initiating evaluative research in treatment facilities for severely emotionally and behaviourally disturbed children. These issues need to be recognised and promptly addressed if suitable research methods and associated procedures are to be put in place. Yet, these issues are frequently misrepresented in the literature. There are purported difficulties which on closer inspection prove to be confusions around prescribed research designs and methods, such as those outlined in the previous section. At the same time, other important pragmatic issues in these settings remain less readily discussed. These include essential issues which need to be addressed if the study design is to prove sufficiently robust to survive in the setting to which it is to be introduced, whilst maintaining the integrity of the research program.

As noted in Chapter 1, there is little doubt that the current “vagaries of service evolution in the field of residential treatment and ideological dilemmas intrinsic to institutional care of children and adolescents” (Parry-Jones, 1998, p.3) has placed the viability of institutional facilities “on the crossroads” and in “crisis”. Yet, as argued on the basis of recent publications, the responsiveness of these units to the challenge of demonstrating effectiveness or even “evidence to guide practice” is not very encouraging. In the following, two recent publications which specifically outline pragmatic difficulties in conducting evaluative research in these settings are examined. This leads to a discussion of the important pragmatic issues which remain unaddressed in the publications.

2.31 Some purported difficulties

Two relatively recent publications will be discussed.
The first publication to be discussed is *Child Psychiatric Units at the Crossroads* (1996), edited by Rosemary Chesson and Douglas Chisholm. This book is mainly concerned with describing in depth the work of the Lowit Unit at the Royal Aberdeen Children’s Hospital, Scotland. Both book editors and almost all of the chapter authors are identified as employees or former employees of the Hospital.

As mentioned in Chapter 1 (General Introduction), the editors show clear recognition from the outset that “unless arguments for units’ survival are supported by detailed, up-to-date information on their functions, treatment processes and outcomes, then purchasers may not be prepared to support these seemingly costly facilities” (from the preface, p. ix). However, the only chapter on research in the Unit (*Dimensions of therapeutic input: a research study*, Chesson, 1996) describes a qualitative study, and the section within it pertaining to treatment outcomes records staff views with regard to their own subjective impressions of their effectiveness. The quoted staff comments touch on the following: “beliefs” that the personal investment they make will have an effect “even if it’s in five years time”; their need to consider their achievements in the context of “the most disturbed families” where improvement takes place “up the ladder one or two rungs - we are not able to change things in a major way”; the non-suitability of some children and their families to treatment, resulting in effective help for only “some children”; and the “widely recognised” notion that “outcomes were not easily quantifiable” as, for example, “forming relationships with the children” is “important regardless of outcome” (p. 215 – 216).

Many of these sentiments are readily recognisable to those with experience in therapeutic units that work intensively with emotionally disturbed children. The quoted comment that forming
relationships with troubled children is important “regardless of outcome” is particularly poignant in the present context – in the McCord (1978) study (discussed above [2.22]) there appears to have been little doubt among the youths, or their counsellors, that the work in which they were engaging was valuable and rewarding. This raises an important question about whether the enormous efforts sustained in working with troubled youths, often with little material or personal gain, can only be sustained through an illusion – a belief with no basis in fact – that one is engaged in important work. It also raises the related question of whether “forming relationships with the children” seems important to staff regardless of outcome because it meets a need in staff. This in turn raises the question of whether staff may be providing a service on behalf of the need of the broader community to feel something of value is in place for troubled children and youths.

Chesson’s research chapter does not enter into a discussion of what instruments or indices are necessary to detect an effect “in five years time”; nor how to detect improvement of only “one or two rungs” in very disturbed children; nor what designs are necessary to test staff hypotheses about the non-suitability of some children and their families, the predicted better outcomes for the more suitable subgroup, and the characteristics which differentiate the two subgroups; nor the means by which “important” outcomes might be assessed. Their qualitative study stands isolated from the large literature on all of these methodological issues, and falls strikingly short of the needs which the book sets out to meet. That is, the book offers no evidence of objective outcomes at the Lowit Unit which might reasonably inform their “purchasers” of what they can expect to get for their money.

The second of the two recent publications to be discussed in this section is *In-patient Child Psychiatry: Modern Practice, Research and the Future*, edited by Jonathan Green and Brian
Jacobs (1998a). This book places “research” prominently in its subtitle and unambiguously declares its aim to “produce a book that reflects the move towards evidence based medicine by emphasising the links between research and practice and, wherever possible, the evidence base for that practice” (from the introduction, p.1). However, the part of the book devoted to “Research” - Part VI - consists of two short chapters and is at 14 pages the shortest Part by some margin.

The first of these chapters, Research into efficacy and process of treatment (Imrie & Green, 1998, pp. 333 - 338), states that it will examine “studies of efficacy” (that is, the question of whether inpatient treatment for children is in fact effective) yet proceeds to provide an uncritical summary of the conclusions by Pfeiffer & Strzelecki (1990) with regard to predictors of outcome (that is, the factors accounting for variation in outcomes). The chapter concludes that “the slow development of this field is related”, among other things, “to the considerable methodological problems in pursuing the area” (p. 337). There is no recognition that the appropriate research design is quite different for these two different types of questions, as has been pointed out in the previous section [2.2].

The second of these two chapters, Methodological issues and future directions for in-patient research (Green & Jacobs, 1998b, pp. 339 – 346), elaborates these “considerable methodological problems”. They are illustrated through an account of one of the authors’ (Jacobs) experience of attempting to maintain adherence to a research protocol across four London child inpatient units studying the effectiveness of a new parent management training program for “6 -9 year old boys with severe mixed disorders of conduct and emotion” (p. 339). These reported difficulties include the following:
“One unit had to withdraw because they did not have sufficient staff to deliver the treatment programme in addition to their usual programmes. A second unit was, surprisingly, unable to recruit suitable patients during the trial period… As the therapy phase of the trial came to an end, one of the two [remaining] units providing the treatment was suddenly, and inappropriately, closed!” (p. 339 – 340).

The authors also describe the extra clinical resources necessary to sustain the new parent management training program, the difficulties for the units of admitting children in accordance with the research protocol due to the “constraints of difficult youngsters already admitted” (p. 340)\(^4\), and the difficulties to families of participation in the program “unless their child care and travel costs can be refunded” (p. 340).

The authors appear to present these experiences as an illustration of best research intentions being defeated by a series of difficulties beyond researchers’ control. While they acknowledge that “clinical staff perceive research (especially efficacy research) as potentially threatening to their self-confidence and unit culture” and that the implicit “possibility of change in clinical practice” may be resisted by “some unit cultures” (p. 341), the authors appear to understate the pervasiveness and strength of the resistance to change. They also leave unexplored its likely sources. They appear to be unaware of the inadequacy of the described research procedures in the face of this resistance, and also of the provocativeness of an initiative such as theirs (i.e. the introduction of a new treatment and its associated procedures) in their setting. Furthermore, their

\(^4\) I take this to be a reference to the usual considerations in the timing of a new admission to a child inpatient unit with regard to its likely impact on the milieu. These might involve balancing, on one hand, the urgency of the case against, on the other, the factors which determine the capacity of the unit at that time to contain the newly admitted youth (such as the level of distress in the unit, the mixture of children currently in admission, bed numbers, and changes in staff availability due to leave, transfer, rotation, holidays, and so forth).
discussion of methodological difficulties and potential solutions suggest a pre-commitment to particular research methods which are quite inappropriate.

These important issues, which tend to be neglected in the literature, require more careful consideration.

2.3.2 Some neglected issues

Three neglected issues will be discussed in turn: the “resistance to change” of organisational procedures; the uncritical adherence to “prescribed” research methods; and the resistance to outcome evaluation located within the researchers.

The “resistance to change” of organisational procedures: The inevitable conflict between the interests of administrators who wish to present their programmes in the best light and the interests of researchers who evaluate them is well recognised in the social sciences (Campbell, 1969). Furthermore, there are theoretically well-worked out approaches to understanding and containing anxiety in institutions, including the anxiety inevitably associated with change (Menzies, 1959; Bain, 1998, 1999; Long & Newton, 1997). An intervention to a social system which needs to manage very intense, conflicting emotions on behalf of the broader society - such as hospitals, the military, mental health facilities, emergency services, the justice system, gaols, and so forth – is unlikely to succeed without an understanding of the way in which its procedures and organization, ineffective though they may appear with regard to its purported tasks, might be serving to protect the individuals who work in them from too direct and intense exposure to distressing experiences.
For example, the “red tape” researchers may perceive as an obstacle around admitting children according to their research protocol is likely to be important in providing clinical staff with a sense of control over the intense milieu and the case-load they are carrying at any time. In this sense, unit procedures may operate as a “social defence” (Menzies, 1959) against overwhelming distress and anxieties. Social defences can be expected to be prominent in agencies with the task of managing the most distressed and anxious children and families (and hence some of the most distressing and anxiety-provoking) in the community. Whether these are mature and flexible defences (operating in units which are on-task and well-functioning) or rigid and ineffective defences (operating in units which are off-task and poorly functioning), it should not be surprising that researchers who attempt to cut across them discover to their peril that their research protocol is met with open hostility, given a low priority, or undermined by the countless number of passive-aggressive actions possible in any large organisation leading to the familiar “death by a thousand cuts”.

Uncritical adherence to “prescribed” research methods: Green and Jacobs state the following at the outset of their chapter on methodological problems in conducting evaluative research in their setting.

“The problems in trying to carry out research in child psychiatry in-patient units are significant. Because there is a very low incidence of major psychiatric disorder within this age range the strategies used for the most part in adult services do not apply. This is compounded by the length of stay of children in in-patient units, again rather longer than adult services, leading to a low throughput and the slow accumulation of any series of patients.

In fairness, it needs to be admitted that the most distressed (and distressing) youths are probably to be
Further, the nature of the cases is heterogenous. Most cases show multiple comorbidity, so that the development of clear populations with few variables is improbable. Inevitably this leads to the need for larger populations to compare and analyse. To complicate matters further, the culture on in-patient units is not stable. It fluctuates, as discussed elsewhere in this volume (chapters 29, 2). This should not produce a difficulty for research provided that the culture fluctuates about a mean. However, it is at least possible that there are features of the culture and practice on the unit that change progressively over time. Methods for recording these matters so that they may be factored into research efforts are still in development.” (1998, p. 339)

This well illustrates the unhelpful adherence to notions of “prescribed methods” discussed in the previous section [2.2]. It might be asked of the authors why it is important to conduct research on the small number of cases which fall within current diagnostic categories of “major psychiatric disorder” and without “comorbidity”, rather than the children with various emotional and behavioural problems that are actually admitted to their unit. Furthermore, if one is not attempting to accumulate a large-enough homogenous subsample, the longer admissions compared to adult inpatient settings might be considered an advantage rather than a disadvantage. A longer admission allows, for example, the evaluation of treatments not able to be offered in the “revolving door”, as it has been called, of brief emergency admission, stabilisation through heavy doses of medication, then discharge to community management services, which characterises adult mental health services in the swing away from “institutionalisation” in the Anglosphere since the 1960s (Cohen, 1988).

found not in the mental health facilities but the juvenile justice facilities.
With regard to recording changes in unit culture over time, the authors provide no indicators as to what sorts of methods “in development” these might be. An approach the authors do not consider is to, first of all, recognise that it can be confidently expected that change in unit culture will happen, and secondly, to consider this inevitable change an object of study rather than random error – that is, “signal” worthy of attention rather than random “noise” which might be hoped to be statistically cancelled out. Changes in the unit culture are likely to be associated with cycles around, for example, the departure and arrival of unit directors and other influential individuals; scandals or crises; changes in the parent body to which the unit is answerable; changes in government policy; or broader trends such as the current swing away from institutionalisation or the swing towards “empirically-based” treatment. While it will not be possible to anticipate what sort of change these might be, it is very likely that the changes in culture can be readily identified retrospectively. With clothing fashions, for example, while it may not be possible to anticipate its vicissitudes, it is not difficult to retrospectively recognise periods when men’s neckties grew thinner or wider, or the wearing of business head wear waned, and so forth. (Retrospective classification of changes in unit culture is an important feature of the present study, described in Chapters 4, 5 and 6).

Similarly, the provocative and divisive clash between admitting children in accordance with an introduced “research protocol”, on one hand, and admitting them according to the “usual” clinical and pragmatic criteria, on the other, can be avoided by recruiting “research subjects” from among those already admitted. (This is another feature of the present study, described in Chapters 4, 5 and 6).
With regard to the “difficult” task of maintaining “uniform treatment across different sites” (p. 340), the authors do not explore the alternative of allowing the natural differences between units to remain and then examining the differences in outcomes associated with these differences (i.e. a quasi-experimental design).

**Resistance to outcome evaluation located within the researchers:** A pre-commitment to particular research methods which require a change in the operation of the Unit, evident in Green & Jacobs (1998b), rather than the initiation of methods which provide information about the actual operation of the Unit, as suggested above, may reflect the researcher’s own resistance to outcome evaluation. While locating resistance to research in “some unit cultures”, the authors fail to acknowledge that their interest in the effectiveness of the new treatment involves turning away from the question of the effectiveness of the existing treatment in their unit. It is difficult to estimate whether an explicit inquiry into the effectiveness of a unit functioning as is is likely to produce more anxiety and resistance than the evaluation of a new treatment. In either case, as mentioned, the authors’ cursory treatment of the subject of resistance to research appears to serve to demonstrate the difficulties faced in fulfilling their stated aim to “produce a book that reflects the move towards evidence based medicine by emphasising the links between research and practice” (p. 1) rather than take a serious step towards this aim.

Moreover, the limited range of potential solutions offered by the authors, and their lack of interest in studying the actual present functioning of their Unit, show how it is possible to resist inquiry while claiming to be adhering to “scientific methods”. This is an example of how the call for “evidence” and “evidence-based practice” can itself serve as a defence against the collection and examination of evidence.
2.33 Anxieties underlying resistance to evaluation

It might be asked, then, whether there are special conditions in inpatient child units which produce such pervasiveness and strength of resistance to evaluative research. This subsection, quite speculative in parts, attempts to address this question.

The methodological confusions - outlined in the previous section [2.2] - are by no means unique to this area. Nor are the pragmatic difficulties – as outlined in this section [2.3] - unique to this area. It needs to be asked why there are so few reports on the types of studies which might be expected to be both manageable and useful in this type of setting. These studies might involve, for example, an examination of the predictors of outcome for the children entering these units, or comparisons of the effectiveness of these units with less intensive, less costly services.

It is a fundamental principal of psychodynamic theory that defences come into place in response to anxiety associated with an unmanageable or unrecognised affect. The clearly maladaptive paralysis in inpatient child units – a failure to produce evidence of its effectiveness, or even evidence to guide practice, despite conscious recognition of the consequences of continued inaction – was introduced in Chapter 1 and highlighted again at the beginning of this section [2.3]. The examination of two recent books offered evidence of defensive, avoidant behaviours in this respect [2.32]. It might be asked then, what is the anxiety, and what is the avoided affect which provokes this anxiety?

It appears that the most superficial level of resistance to research, the clerical and the administrative resistance to the collection of data, is most readily discussed in the recent literature. The staff involved in the day to day administrative duties, usually with longer (and
perhaps more secure) tenure than the clinical staff or unit director, correctly believe they run the Units. They are personally identified with the procedures they have developed - or inherited - of admission forms, discharge forms and regulatory statistical reports, as well as staff rosters, timetables and access to material resources. It is more efficient, therefore, to access the information routinely collected or, alternatively, to influence the information routinely collected, than to introduce additional, competing procedures. Green and Jacobs (1998) recognise this in their reference to recommendations by Moore & O’Connor (1991) and Gerardot, Thyer, Mabe & Poston (1992) that “a standardised component of evaluative research be built into normal clinical practice” (p. 341). This can be taken much further, as argued above; the research protocol can be brought into further alignment with the everyday clinical operations of the Unit by studying the population of youths actually admitted rather than admitting a different population, or admitting them differently for the purposes of study. (This is an important feature of the present study, described in Chapters 4, 5 and 6).

However, a more complex and much less tractable level of resistance arises from anxieties about the way in which the data might be used. This can ignite conflicting interests within the Unit, and also between the Unit and the agency to which it is accountable. This is not a subject readily explored in the literature.

With regard to conflicting interests within the unit, health workers are rightly suspicious of negative aspects of large-group dynamics being exploited by managers to achieve their aims. This is unlikely to cease while public institutions fail to provide managers with clear leadership accountability and the means by which they can legitimately discharge that accountability (cf. Jaques & Clement, 1994). These large-group dynamics might include scape-goating and paranoid anxieties, often associated with the question of responsibility (i.e. “who to blame”) for
an incident or a crisis, or seizing at a partial and simplistic explanation of complex situations, such as self-serving interpretations of health statistics in the mass media.

The way in which the issue of how the data is used inevitably touches on the powerful issue of who is being evaluated. The patient or the unit? The unit administrator or the individual clinician? When these anxieties are not able to be adequately contained in a Unit, it can be expected that the research program becomes unviable. In this way, the same evaluative research program “on paper” can be conducted in a spirit of inquiry or become inhabited with paranoia, cynicism and distrust, according to the social dynamics of the institution and the way in which it is managed.

Similar issues arise with regard to conflicting interests between the Unit and the agencies to which it is administratively accountable, typically hospital management, university departments or government departments. A special case is the tension between professional groups aligned with universities, on one hand, and professional groups aligned with treatment centres, on the other. This tension is very prominent in the recent literature on “empirically supported” psychological treatment - for example, the Persons & Silberschatz (1998) debate discussed above [2.22], which appeared in the special section on “Empirically Supported Therapies” in the Journal of Consulting and Clinical Psychology. The former group is typically interested in conducting well-controlled clinical trials, thereby determining which specific, defined treatment types are effective with which specific, defined patient groups. The latter group typically claim that what may show “effectiveness” in RCTs conducted in special settings may not show “efficacy” in general clinical settings. The confusion evident in the language⁶ suggests that clear

---

⁶ There appears to be a preference (cf. footnote 4, page 14, Chambless & Hollon, 1998) for framing this discussion in terms of “efficacy” and “effectiveness” over more clearly defined terms such as internal and external validity. Dictionary meanings of both words refer to the ability to produce an effect.
thinking is difficult when acrimonious debates are conducted without articulation of the underlying conflict of interest - academics keen to prescribe scientific and clinical practice, on one hand, and clinicians resentful and fearful of any change in work practices being imposed upon them, on the other. The difficulties were not helped by the emergence of the entrepreneur-professor, increasingly prominent in Australian universities in the 1990s with the swing away from objective inquiry towards credentialism and professional interests (Michell, 1996), of whom the health worker is rightly suspicious.

It needs to be admitted that clinicians usually come to identify very strongly with a particular way of working with people, and do not welcome demands for change in this respect. This, of course, is inconsistent with the requirements of empirically-based practice. However, to characterise this personal identification as an unfortunate deviation from the “scientist-practitioner” model presumes a naïve, idealised view of human scientific endeavour. For example, the evidence for the influence of “investigator allegiance” on outcome variability across studies is now quite strong (Luborsky, Singer & Luborsky, 1975; Smith & Glass, 1977); any therapy tends to do better in comparison with rival therapies when the study is conducted by individuals strongly personally identified with that therapy type. Also, it is not always remembered that Kuhn’s (1962) influential notion of paradigm change arose from a historical survey which showed not that accumulation of evidence leads to a change in the views held within a particular field at a point in time, but that the new orthodoxy emerges over time as younger scientists coming into the field readily identify with the “revolutionary” paradigm while those holding to the superseded paradigm simply die out.

However, the negative aspects of large group dynamics which have been described are not particular to health organisations, much less child inpatient units. Nor are concerns that
internally collected data may be used, or abused, to implement unwelcome changes to preferred and familiar work practices. They fail to satisfy as special conditions of inpatient child units which might explain its particular paralysis with regard to evaluative research.

Socio-analytic theory (Bain, 1999), the extension of psychodynamic understanding to social systems, suggests that where the behaviours and achievements of individuals in institutions are significantly at odds with the personal capacity of these individuals, the social dynamics of the institution need to be considered. Furthermore, it suggests that the dynamics which interfere with the institution’s primary task (i.e. that which needs to be accomplished for the institution to survive) are usually associated with difficult emotional experiences arising from the work in which the institution is engaged. More specifically, these experiences and associated dynamics are likely to be found (a) within the “clients” of the organisation, (b) within the organisation itself, and (c) in the relations between the organisation and its clients. These theoretical considerations provide a way of approaching the question of what are the unmanaged or unrecognised emotional experiences in the Units interfering with the self-evaluative function.

In this regard, it is noted that both Chesson & Chisholm (1996) and Green & Jacobs (1998) – the two publications considered in some detail in the previous subsections [2.31, 2.32] - contain chapters concerned with the internal dynamics of the units. These include the perspectives of multi-disciplinary staff with regard to their management of their relationship with emotionally disturbed youth, and with regard to their management of the relationship between themselves. Yet, there is very little attention to the dynamics of the external relations between the units and the broader community. This imbalance, in fact, appears to reflect the general literature on institutional treatment of young people. For example, the imbalance is present even in the otherwise very detailed and comprehensive volume describing the clinical work of the Cassel
Hospital, South West London, a leader in the inpatient treatment of families and adolescents within a therapeutic community (Kennedy, Heymans and Tischler, 1986). This is in contrast to the broader literature available on the social dynamics of adult psychiatric hospitals (e.g. Goffman, 1961, 1969; Scott, 1973; Scott & Alwyn, 1978; Menzies Lyth, 1989; Bott, 1976). This contrast may be significant; it may reflect an unexamined aspect of the hospital treatment of children.

It may be productive, then, to consider the more complete understanding of the function of adult institutions. Some results of the difficult work of bringing together the various strands of understanding of the internal and external relations of the adult psychiatric hospital can be found in Bott (1976, 1990). Bott describes the dominant “schizophrenic” quality of the internal relations of mental hospitals. She quotes a young doctor working in a British mental hospital in 1957.

"The hospital is like schizophrenia itself. Split up in bits, projections all over the place, parts not communicating with other parts. Things are always getting lost in this place - people, ideas, decisions. There is an overpowering sense of inertia" (1976, p.129).

Bott’s wide-ranging paper finds an analogue of these “schizophrenic” qualities in the family relations of the patient. She relies in this respect on the clinical observations by Scott (Scott, 1973; Scott & Alwyn, 1978) of an abrupt discontinuity from emotional over-involvement with, and idealisation of, the prospective patient by their parents prior to the first admission, to the emotional disengagement and disenchantment that takes place following admission. She also finds an analogue in the external relations of the hospital to the community, where the medical
solution it provides ("diagnosis", "treatment" and "rehabilitation" of the "patient") remains detached from the family process in which it is asked to intervene.

Turning back to inpatient child units, their relations with the community are quite different from adult psychiatric hospitals. At least in the type of unit which is the focus of the present thesis, those offering intensive psycho-social intervention\(^7\), the family is the unit of intervention. If the unstated contract between the adult mental hospital and the family is "your family member is suffering an illness, leave them here and we will treat their illness", the unstated contract between the child unit and the family might be "you have been unsuccessful in raising your child, we will provide better parenting here and also teach you to be better parents". Furthermore, the emotional qualities of a child unit conducted along milieu-therapy lines are far from inert. While ideas and decisions, and even some staff, may get lost in these places, the children do not. On the contrary, there are very intense emotional relations formed with and around the children (MacDonald, 1991; Redl & Wineman, 1951).

It may be the case that the special conditions in inpatient child units which produce resistance to self-evaluation is associated in some way with this intensity of emotional attachment, on one hand, and the unstated contract with the parents, on the other. Operating with such an intense mixture of hope and despair, as reflected in the qualitative research by Chesson (1996) discussed above [2.31], it may be that these Units prefer to keep operating on the basis of phantasies about outcomes – that is, implicit, omnipotent assumptions about their own extra-ordinary parenting competence - rather than subject themselves to the reality test of objective data. The source of anxiety, then, might be associated with an unmanaged and unrecognised fear that the children

\(^7\) As opposed to those offering custodial care or primarily medical-pharmaceutical intervention. Arndell, for example, adhered to a policy of admitting only children and adolescents with identified "parents" (whether they be composed of two-natural parents, single-parents, grand-parents and foster-parents) committed to working with the treatment program and offering a continuity of care following discharge.
are not being helped, that the Unit might be failing the children in their care. This becomes particularly poignant when staff come to believe that they represent “the last chance” for these children.

It can be readily seen that these kinds of staff experiences are analogous to that of parents with whom the Units work. These unwanted feelings are perhaps able to be located in the parents during the period of treatment, but become intolerable once the treatment comes to an end, which would normally mark the commencement of the follow-up phase of evaluative research. The Unit perhaps then comes to be burdened with a phantasy of retaliation for their attitude of superiority to the parents - a fear that they will receive their “comeuppance”. Under such circumstances, an ongoing evaluative program may take on the emotional resonance of a petard of their own making.

It has been argued that the existence of unarticulated, special anxieties is necessary to explain the unusual paralysis of outcome research in specialist facilities for severely emotionally disturbed children. The suggestion is made that intense unrecognised fear of public humiliation and aggressive retaliation may play a more important role in institutions of this type than, for example, in outpatient settings where the adoption by the institution of the role of parent is much less total.

In conducting the project from which the data for the present study, described in Chapters 4, 5 and 6, was extracted, an attempt was made to contain staff and management anxieties by feeding back emerging results to the Unit first, rather than an external body or forum. If it is the case that anxieties associated with phantasies of retaliation and failure, as described, or illusions about the effectiveness of one’s work, as outlined earlier, characterise the emotional life of these
institutions, they clearly present challenges to the conduct of evaluative research program. Equally, an effective evaluative research program serves to contain these anxieties and potentially improve the conduct of the clinical activities.
2.4 Confusion about “clinical significance”

Along with the recent interest in the methodological and conceptual aspects of the evaluation of clinical outcome, there has been increased interest in the difference between “statistical significance” and “clinical significance” (e.g. special edition of the *Journal of Consulting and Clinical Psychology* in June 1999). The discussion of clinical significance and methods for its measurement has become increasingly incorporated into the mainstream of psychotherapy research (e.g. Bergin & Garfield, Handbook, 4th ed). The interest in “clinical significance” appears to reflect a growing appreciation of the limitations of what can be learned about the phenomena of interest – psychological change in individual adults and children – from the conventional reporting of the results of tests of statistical significance performed on pre and post treatment data. This would appear, in part, to be part of the growing criticism of the usual practice of “significance testing” in the social sciences, a criticism which has a long history (Morrison & Henkel, 1970), and may lead editorial policies in mainstream psychology journals towards greater “inclusiveness” (Crassini, 1997).

However, many different issues are touched on in the “clinical significance” debate. While the recent efforts to develop new methods for determining the significance of clinical change has brought welcome attention to neglected issues, it also appears to be perpetuating old confusions and perhaps even introducing additional confusions. For example, the authors represented in the special section on clinical significance in the June 1999 issue of the *Journal of Consulting and Clinical Psychology* clearly mean different things by “clinical significance” and touch on quite a range of different issues, from treatment effect size in the individual case (Jacobson, Roberts, 

---

8 This literature is mainly concerned with the issue of whether the current practice of data analysis can lead to meaningful conclusions about psychological theories (e.g. Meehl, 1991). The more fundamental conceptual issue of scientific inference from data to theory (e.g. Grayson, Pattison & Robins, 1997) is less frequently discussed. The focus of the present section will remain on examining outcome in a technical and pragmatic sense, and not the significance of the outcome with regard to competing theories of psychopathology or treatment.
Berns & McGlinchey, 1999) to the “social acceptability” of treatment goals (Foster & Mash, 1999). While this heterogeneity is acknowledged in both the opening precis (Kendall, 1999) and the concluding commentary (Kazdin, 1999), no attempt to integrate this growing area is offered. Furthermore, there are other important aspects of the clinical significance of outcome that do not emerge in these articles.

It will be argued in this section that (a) the current methods which measure “clinical significance” (versus statistical significance) conflate two separate issues: quantitative aspects of outcome, and the “clinical significance” of these outcomes; and that (b) the judgement of what “clinical significance” any outcome possesses is a qualitative issue. It will then be argued that the recent discussions of “clinical significance” have failed to separate these issues from another important distinction – that between conventional analysis of group data and the analysis of data on a case-by-case basis. A brief history of the confusions associated with “single-case experimental designs” is provided. This section concludes with a discussion of methods for calculating effect size which most readily lends itself to consideration of the clinical significance of treatment outcomes.

2.4.1 Current measures and indices of “clinical significance”

It needs to be stated at the outset that the application of the term “clinical significance” to blatantly subjective methods - such as the general, impressionistic opinion of therapists whether their work has resulted in “significant” change - will not be considered. The vulnerability of such methods to the demand characteristics of the situation are obvious and will not be outlined here.
Yet even after restricting the application of the term in this way, the discussions of clinical significance appear to be struggling to come to terms with quantitative aspects of research results beyond the recognition that the statistical significance of the results is not a measure of the strength or size of the results.

In fact this mistaken notion often still reappears in the literature on treatment outcomes. For this reason, it will be briefly outlined before proceeding to a discussion of current measures and indices of “clinical significance”.

For example, as the $p$ value generated by a statistical test of the difference between two treatment groups is a function of (i) the size of between-group difference, (ii) the size of within-group differences, and (iii) the number of subjects, any difference between groups can be shown to be statistically significant given sufficiently large N. Yet, psychologists seem unable to resist attributing to the specific finding of statistical significance – the probability of obtaining the results at hand if there was zero difference between the groups - the broader, general meanings of “significance”. The “slippery slope”, as Cohen (1965) puts it, of “If it’s statistically significant, it’s important, consequential, worth talking about, large, that is, significant! (p.532)”

Two illustrations of the continuing strength of this confusion in the field will be provided. The first illustration comes from general psychology and the second illustration is more specific to therapy outcome research.

First, while Cohen’s well-known book addressing the neglect of power analysis - and its corollary, the neglect of effect size - was first published in 1969, recent power surveys show
little change in research practice (Sedlmeier & Gigerenzer, 1989). The power of the null hypothesis test remains typically “below .50 for a medium effect” (Gigerenzer, 1993).

Second, summaries of psychotherapy research outcomes often still fail to address the magnitude of the phenomena and merely catalogue studies reporting statistically significant results⁹. This is so even in What Works for Whom: a Critical Review of Psychotherapy Research by Roth and Fonagy (1996) a large and relatively sophisticated survey of the psychotherapy literature which took on the enormous task of determining what empirical support exists for the various psychotherapies with reference to specific populations, symptoms and settings, and which preceded by several years the aforementioned What Works for Whom: a Critical Review of Treatments for Children and Adolescents by Fonagy, Target, Cottrell, Phillips & Kurtz (2002). Roth and Fonagy attend to the difference between “clinical and statistical significance” under “problems associated with the use of statistical tests in psychotherapy research” (p.24).

They provide three different strategies to detect “clinically significant therapeutic change”.

“● Comparison of patient change with normative samples,
● Measurement of the extent of individual change by reference to a criterion measure of change – for example, that treated clients should be 2 SD from the mean of the untreated group (Jacobson & Truax, 1991)

⁹ While obviously not a problem restricted to research in child units, it was not infrequent that papers reviewed for Chapter 3 failed even to report means and standard deviations; they provided t or F values, or p values, or stated that a conventional (.05 or .01) level of statistical significance was achieved. This meant that the interested readers were unable to form their own views about the effect size in a particular study.
The use of criterion of recovery which enables categorical rather than continuous scoring of outcomes – for example, using a BDI score $\leq 9$ to indicate treatment response (e.g. Elkin et al, 1989).” (p. 24)

Furthermore, they state that:

“The clinical significance of change is central to the evaluation of psychotherapy outcomes; though recent investigations are more likely to report data in this form, such measures are not always available” (p.24)

The authors then adopt in their book the usual practice of referring to significant $p$ values as “significant” results.

Following this brief outline of the strength with which the erroneous but appealing ideas about $p$ are held, the appropriate place to commence the discussion of the current measures and indices of “clinical significance” may be the indices of clinical change developed by Jacobson and his colleagues. Its appearance in the first quote from Roth and Fonagy reflects its prominence; they are probably the most widely adopted indices of clinical significance in the recent research literature.

In a series of influential articles (Jacobson, Follette & Revenstorf, 1984; Jacobson & Truax, 1991; Jacobson, Roberts, Berns & McGlinchey, 1999) Jacobson and his colleagues developed a systematic approach to the clinical significance of the results of psychotherapy and behaviour change interventions. They provide two different indices for two different aspects of clinical
significance - the “reliable change index” and the patient’s post-treatment “clinical status”. They are discussed in turn.

Jacobson’s reliable change index (RCI) is concerned with determining whether or not the magnitude of change seen in an individual’s test scores exceeds that which might be attributable to the unreliability of the test instrument. This index has been readily adopted by the psychodynamically oriented (eg. Target & Fonagy, 1994) as well as the behaviourally oriented, and has now become accepted as a standard method.

The RCI is defined as the difference between an individual’s pretest score and posttest score divided by the standard error of the difference between the two test scores.

\[
\text{RELIABLE CHANGE} = \frac{\text{POSTTEST SCORE} - \text{PRETEST SCORE}}{\text{STANDARD ERROR OF THE DIFFERENCE BETWEEN THE TWO TEST SCORES (S_{diff})}}
\]

Standard error of the difference between the two test scores is computed from the standard error of measurement.

\[
S_{\text{diff}} = \sqrt{2(SE)^2}
\]

The standard error of measurement in turn can be computed from the standard deviation of test scores in the normal population and the test-retest reliability.

\[
SE = sd\sqrt{1 - r_{xx}}
\]

Where the index is greater than 1.96 the probability that the change in test score is due to random fluctuation in the measurement instrument is considered to be less than .05 and hence the change seen in the individual is considered reliable (from Jacobson & Truax, 1991).
They provide the example of Spanier’s Dyadic Adjustment Scale, a well known instrument for rating marital satisfaction and dissatisfaction. A reliable change by this criterion requires a pretest to posttest difference of at least 10 points. That is, an increase of 10 points indicates a reliable improvement, a decrease of 10 points indicates a reliable deterioration, and a difference of less than 10 points cannot be regarded as reliable change.

Its appeal over the standard practice of reporting $p$ values, leading to its adoption as a new standard, appears to lie in the following: (a) it is presented as a direct measure of effect size; (b) moreover, it appears to be a measure of effect size in the individual case; and (c) it appears to provide a precise quantitative criterion for reliable change and thereby the reassurance of a seemingly objective mathematical method.

The special appeal of its being a measure of change in the individual case, as opposed to change in the group undergoing treatment, will be discussed in the next subsection [2.42] on confusions around individual (case-by-case) versus group (traditional) models of data analysis and will not be discussed here. This aside, there are several important qualifications about the RCI which need to be spelled out.

First, it needs to be stated that the RCI is not a direct measure of effect size, and that the assumptions underlying the calculation of the RCI for any instrument are quite arbitrary. Care needs to be taken in not overestimating the “significance” of the finding that an individual changed more than is expected by random variation due to “test reliability” – the amount of change required to be considered reliable on a particular instrument does not indicate how “small” or “large” it is. The role that test reliability plays in the calculation of the RCI needs to be spelled out. It can be seen that the RCI is a function of (a) the amount of change seen on an
instrument, (b) the test-retest reliability of the instrument, and (c) the selected level of reliability (usually .05). Just as any difference between groups can be shown to be statistically significant given sufficiently large N, any positive pretest posttest change in an individual will be a “reliable change” given a sufficiently reliable instrument. Therefore, while it may be deemed desirable to select instruments with increasing reliability to increase sensitivity to small changes, this will have the effect of a “positive” RCI being associated with increasingly trivial magnitudes of change. Conversely, a positive RCI on a relatively insensitive instrument will be associated with a larger magnitude of change.

For example, the Child Behaviour Checklist (CBCL) completed by parents has higher reported test-retest reliability than the parallel instrument for teachers, the Teacher Report Form (TRF). Should a larger proportion of subjects, following a treatment intervention, show reliable change on the CBCL than on the TRF, it would be erroneous to conclude that the intervention showed more home based change than school based change. Due to differences in the reliability of the instruments, and hence the associated sensitivity of the RCI, a greater change would need to be shown in the ratings by teachers using the TRF than the ratings by parents using the CBCL to register a positive RCI. This means that while the RCI facilitates some types of comparisons between different populations under study that use the same instrument (i.e. the proportion of subjects showing reliable improvement on a particular instrument calibrated according to a particular test-retest reliability), it is not readily applicable to comparisons involving different instruments. For example, Jacobson, Follette, Revenstorf, Baucom, Hahlweig and Margolin (1984) compared four different outcome studies of behavioural marital therapy with regard to the proportion of couples showing improvement on the RCI, despite three different instruments having been used!
Second, it needs to be noted that the RCI is a three way categorisation of pretest to posttest change (reliable improvement, reliable deterioration, or no reliable change) in an individual and not a continuous measure of the amount of change. The loss of information involved in the compression of data (i.e. where individuals with the minimum level of pretest to posttest change to satisfy the criterion are assigned to the same category as those with much larger changes) means that conclusions about effectiveness need to be restricted to statements about the proportion of subjects making a defined minimal change. Such a data method is not well suited to research questions involving, for example, a comparison of treatment responses across subjects where consideration of the full range of responses is important. It will be important in many research applications to be able to identify the outstandingly successful and unsuccessful cases.

Third, the period over which test-retest reliability (utilised for deriving the standard error of the difference between the two test scores) is calculated will in turn have an impact on the calculation of the RCI. For example, is it appropriate to calculate RCI based on test-retest reliability over one period and apply it to a study of change over a much longer period? It might also be asked whether it is appropriate to utilise test-reliability calculated for a general population and apply it to study on a clinical population, when the latter might perhaps be expected to show greater variation in scores over time? In practice, the arbitrariness in obtaining a standard error for the purposes of calculating RCI extends to substituting test-retest reliability with inter-judge reliability without comment (Fonagy & Target, 1994), or using measures of internal consistency instead because test-retest reliability was not available (Jacobson, Follette, Revenstorf, Baucom, Hahlweig and Margolin, 1984, page 499, footnote).
On the basis of these three qualifications, and the evidence of current practice, it is concluded that there is a danger, as with $p$ values, that researchers (a) overinvest reported RCI results with greater, more specific, more objective “significance” than is warranted, and (b) are already applying the RCI in an inappropriate manner.

The second of Jacobson’s two indices of clinical significance is concerned with the level of functioning attained by the individual following treatment. Jacobson and his colleagues are concerned here with whether the individual can or cannot be considered “normal” following treatment. They explore several methods for operationalising this criterion for any instrument (Jacobson & Truax, 1991), all pertaining to the establishment of a cut off score which divides the “functional” from the “dysfunctional” populations. In their view, an individual cannot be considered to have made a “clinically significant” improvement unless their post-treatment functioning places them in the “functional” population. Later, Jacobson more clearly specifies that he does not consider that a clinically significant change can be said to have taken place unless an individual fulfils both criteria: a reliable amount of change from pretest to posttest (i.e. a positive RCI) as well as a posttest score which crosses the cut off score from the dysfunctional to the functional population (Jacobson, Roberts, Berns & McGlinchey, 1999).

There will be no argument with the notion that it is important to differentiate between (a) how much an individual has changed and (b) their level of functioning following treatment. (The confusion arising from failure to maintain distinctions between these two logically different types of “outcome” is discussed in the next section [2.5] of this chapter). Yet, to attribute the general term of “clinical significance” to the level of post-treatment functioning is unhelpful and confusing. The amount of change achieved (Jacobson’s first criterion) is at least as clinically meaningful as the level of functioning achieved (Jacobson’s second criterion). Furthermore, to
attribute the general term of clinical significance to those who fulfil both of Jacobson’s criteria (i.e. his later view) hijacks the term for use with particular clinical populations. For example, Jacobson’s view, if widely adopted, will codify the individual who makes a very large change from extreme dysfunction to a much more moderate dysfunction as not having achieved “clinically significant” improvement. Yet, the individual with much less extreme dysfunction who makes a smaller change but is now placed within the normal range would be regarded as having achieved “clinically significant” improvement. To consider only the latter individual as showing “clinically significant” improvement is confused and unhelpful. Therefore, for clarity of discussion, Jackson’s second criterion will be referred to as post-treatment clinical status – a consideration of how well the patient is doing after receiving treatment when compared to the normal population.

Furthermore, there are aspects of the clinical significance of outcomes which are not addressed by either of Jacobson’s criteria. Three such aspects will be delineated.

First, there are aspects of the clinical significance of individual change which emerge only in the broader context of the individual. For example, an assessment of the size of pre to post treatment change in the context of variability and duration in the symptom over the course of an individual’s life has been emphasised by Malan (1980). It is an important aspect of his method for the evaluation of outcomes of short to medium term psychodynamic psychotherapy. He refers to the finding, in his second series of thirty patients treated at the Tavistock Clinic, of groups of patients with long-term difficulties (some greater than 10 years) who showed a sudden improvement during the course of relatively brief therapy (some as brief as ten sessions) and who maintained the improvement at a rather intensive follow-up several years afterwards. He asks, “It is surely straining credulity to suggest that the improvement ‘just happened’ to occur at
this point” (p. 334). He calculates the ratio of the duration of disturbance to the duration of therapy before the disturbance improved, which he calls the “duration ratio”. He reports that “careful examination of the thirty cases in our second series reveals that there were eleven in which the duration ratio, very conservatively estimated, was greater than ten” (op.cit.). The very detailed account of the change in the patient and the duration of their difficulties which he provides (Malan, 1976a, 1976b) include many instances where a symptom showing little variation over a life-time responded quickly to the commencement of treatment.

Second, another example of clinical significance arising from broader individual context is the meaning of the observed change following treatment; it is not necessarily obvious whether a reliable change in a particular direction by Jacobson’s first criterion represents “improvement” or “deterioration” for an individual! We might consider Megargee’s (1966) identification of two distinct types among youths incarcerated following acts of extreme violence - the undercontrolled youth whose assault forms part of a long history of acting out and impulsiveness, and the overcontrolled youth whose actions are seen to be “out of character”, where inhibited and unassertive boys respond disproportionately to a provocation. As Megargee pointed out, the therapeutic aims for each subtype are quite distinct; increased aggressive behaviour in a youth of the first subtype may indicate a further deterioration in poor frustration tolerance, while for a youth of the second clinical subtype it may indicate a therapeutically desirable increase in assertiveness.

Third, another aspect of the clinical significance of outcome pertains to the significance of the variable itself. Rosenthal (1993) provides the illustration of the early termination of a randomised double blind experiment on the effects of aspirin on reducing heart attacks (Steering Committee of the Physicians’ Health Study Research Group, 1988) when it became “so clear
that aspirin prevented heart attacks (and deaths from heart attacks), that it would be unethical to continue to give half the physician research subjects a placebo” (p. 538). He notes that the magnitude of the experimental effect could be calculated to be an $r^2$ of .0011, with a corresponding $r$ of .034, and that while this is not a magnitude which might generally be considered by behavioural researchers to be very large, the switch to aspirin might be expected to result in more than 3 per 100 who manage to survive. This illustrates the role of robust research designs (randomised) and reliable measurement not usually available to behavioural researchers (whether a subject is dead or alive), in detecting relatively small effects. It also illustrates a point made by Cohen - “Death tends to concentrate the mind. But this in turn reinforces the principle that the size of an effect can only be appraised in the context of the substantive issues involved” (Cohen, 1988, p.534). In this example, not only is the outcome of undoubted importance, the context meant that the interest in medical research gave way to the stronger medical imperative to save lives.

It is argued that there are many different aspects to the clinical significance of outcome results, and the codification of any one (or several) aspect as definitive of “clinical significance” is confused and unhelpful. It represents another example of “prescribed” research method which is the antithesis of the critical inquiry which marks scientific activity.

2.42 Conventional versus single-case data analysis

Another confusion found in recent discussions of “clinical significance” pertains to confusions around the distinction between conventional analysis of group data and the analysis of data on a case-by-case basis. As suggested above [2.41], part of the appeal of Jacobson’s two indices for
those interested in clinical psychology appears to lie in its applicability to individual cases – it allows seemingly objective, mathematical, statistically valid statements about the individual case.

The averaging procedure inherent in conventional group statistics, through which the behaviour of particular variables under study emerge, has as its corollary the neglect of other aspects of variation. This state of affairs may be particularly pertinent to clinical applications. First of all, the calculation of "average" change can obscure the amount of change actually occurring in a group of patients. If some individuals in a treatment group show improvement and others show deterioration, they will tend to cancel out. Secondly, grouping of data in this way necessarily removes a datum from the broader clinical context of the individual case in which it occurs. (Examples have been provided above of the significance of treatment change which can only be appreciated in relation to the individual context [2.41]).

The problems associated with an unsophisticated application of group statistics has long been pointed out from a minority position within clinical psychology. Kiesler (1966), for example, has questioned the assumption of homogeneity of patient response to treatment underlying the group model of data analysis. Chassan (1979) articulated the differences between conventional group-based model of data analysis, which he called extensive data analysis, and individually-based models of data analysis, where clinical improvement is determined on a case by case basis, which he called intensive data analysis. While Chassan’s terms have not been widely adopted in the literature, they will be used at present for the purposes of clarity.

While the issue of intensive and extensive data analysis - two different models of data analysis - is a separate one from the issue of clinical versus statistical significance, and separate again
from the issue of research design, the three issues have become entangled in the development of
the paradigm of the “single-case experiment” (Jayaratne & Levy, 1979; Barlow, Hayes &
Nelson, 1984; Barlow & Hersen, 1984). While these authors brought clarity to the
aforementioned potential limitations of averaging in clinical and applied situations, and
produced an innovation in research design, they introduced confusions which have since become
codified in the current prescriptions of good research practice. The confusion between model of
data analysis and research design arises from their discussion of intensive data analysis with
reference only to a research design based on the systematic introduction and withdrawal of
treatment variables. Despite claims to the contrary\(^\text{10}\), such a design precludes the study of all but
the simplest psycho-social variables, as complex, reciprocal interactions between two or more
individuals are not reducible to the standardised activities of one party; nor can they be brought
under the control of systematic introduction and withdrawal. For example, it might be expected
that the quality of the human relationships (Bowlby, 1988) available to young people would
have an enormous impact within any institutional environment, yet any attempt to bring the
interactions between staff and patients under control through standardisation or systematic
introduction and withdrawal would tend to destroy what natural complexity and spontaneity
they possess. The internal validity of the single-case experiment – the ability to draw valid
causal inferences – rests on experimental manipulation, and this may not be possible with the
variables within an institutional setting, or anywhere else, which may be among the most potent
for patient change.

As for the distinction between the issue of models of data analysis and the issue of clinical
significance, it needs to be stated that there are no necessary connections, for example, between

\(^{10}\) As mentioned earlier [2.22] Persons (in Persons & Silberschatz, 1998), for example, states that single-case experiments are “orientation-neutral” (p. 128). This repeats earlier erroneous claims by, for example, Jayaratne and Levy (1979).
intensive models of data analysis and measures of clinical significance, on one hand, and extensive models of data analysis and \( p \) values, on the other. For example, in the aforementioned special section on clinical significance in the June 1999 issue of the *Journal of Consulting and Clinical Psychology*, Kendall, Marrs-Garcia, Nath & Sheldrick (1999) present a procedure for comparing group outcomes to normative data; it is a measure of a particular aspect of clinical significance which covers very similar territory to procedures by Jacobson, Roberts, Berns & McGlinchey (1999) for comparing *individual* outcomes to normative data\(^\text{11}\). To take a further example, the intensive models of data analysis is not restricted to the visual inspection of graphed time series data to be found in the primers of the single-case experiment; statistically based time series analyses which rely on the calculation of \( p \) values are readily available in many popular statistical software packages.

It needs also to be pointed out that despite the familiarity of the “single-case experimental design”, there is no necessary association between model of data analysis, research design and \( N \). For example, the analysis of outcomes for an experimental study involving many participants might involve an intensive analysis that places each case in one of several distinct outcome categories (eg. improved, no change, deteriorated); the comparison between treatment and control groups can take place on the basis of these categories without any calculation of average outcomes.

The unsound notion that there are particular, “prescribed” research methods to which scientific inquiry needs to adhere has been discussed earlier [2.2, 2.32]. Similar confusions appear to be evident in the special regard with which single-case experimental designs are held in the current prescriptions of methods for conducting psychotherapy outcome research. Single-case

\(^\text{11}\) Although there appears to be no acknowledgement of this by the authors.
experiments were considered “logical equivalents” of randomised clinical trials by the American Psychological Association Task Force on Psychological Intervention Guidelines (Chambless & Hollon, 1998, p. 14). In what amounts to the same thing, the key selection criteria for the survey of child and adolescent treatments by Fonagy, Target, Cottrell, Phillips & Kurtz (2002) include “group design or an experimental single-case design” (p. 39).

While there is no argument with such criteria as a way to cull the number of studies which need to be considered in any broad survey of the field, the uncritical prescription of single-case experiments means the promotion of a method suitable only for a narrow range of treatment variables. It is argued that such prescriptions are contrary to scientific inquiry, and the development of different, more sophisticated methods for more sophisticated treatments.

2.43 “Pure” indices of effect size

The interest in effect size, an essential part of the interest in clinical versus statistical significance, has led to proposals for a “pure” index of effect size - one which is independent of the raw unit of outcome measurement - to facilitate comparisons between studies. Cohen has suggested that the use of “arbitrary measurement units” in behavioural science has contributed to the “lack of a high degree of awareness of the magnitude of phenomena” – “we rarely find ourselves dealing with dollars, years, centimetres or bushels” (1988, p. 532). His proposal (Cohen, 1965) of the use of “pure”, “dimensionless”, “unit–free” measures of effect size such as the common statistics of $r$ (the product moment correlation coefficient) or $r^2$ (proportion of variance), and a convention for defining “small”, “medium” and “large” effect size (where $r^2$ is .01, .09 and .25), has been very influential. This approach has led, for example, to the emergence of meta-analysis (eg. Glass, 1976; Glass, McGaw & Smith, 1981) which seeks to
combine the results from different studies to arrive at a representative, over-all effect size to summarise an entire field of study. It has also led, for example, to the “binomial effect size display” (e.g. Rosenthal & Rubin 1982; Rosenthal, 1993), which represents the treatment effect as differences in outcome rates between the treatment and the control group expressed as a bivariate correlation ($r$).

While this represents a development over “tabular asterisks” (Meehl, 1978) with regard to summarising the body of work in an area, the limitations of this method also need to be recognised.

First, the effect size of a study will be affected by measurement error, with the size of measurement error varying according to the instruments used, the research design, the range of outcomes achieved, and so forth, so that two studies of exactly the same effect may produce quite different $r$ or $r^2$. Far from what may be expected of a “pure” index, the effect size of a study calculated in this way is readily subject to manipulation by the researcher. For example, if effect size is calculated as the difference between the mean pre-test score and the mean post-test score divided by the standard deviation of pre-test scores, the process of subject selection - on ostensibly clinically legitimate grounds such as suitability for treatment - can serve to minimise the pre-test standard deviation, and thereby enhance effect size. As with the RCI discussed above, a measure based on the parameters of the study sample, rather than agreed population parameters, will always be arbitrary.

Second, any systematic loss of context or any systematic obscuring of change associated with a particular model of data analysis (as discussed in the preceding section in relation to the calculation of average outcomes [2,42]) is compounded by further aggregation. Furthermore,
this aggregation may mask important differences between different studies and their respective populations. While this is of course an inevitable part of the inferential process – moving from the individual case to general findings - it needs to be noted that there may be threats to inquiry in the climate of “empirically validated treatment” when whole lines of research or treatment are deemed “legitimate”, or otherwise, on the seeming weight of a meta-analytic conclusion based on a large number of previous studies. Phenomena neglected in the original research studies will not make their way to any meta-analytic summary, no matter how large the number of studies or their subject samples. There is the danger of a self-serving circularity, where the methods for seeking evidence is determined by the research and treatment paradigms it supports, rather like the old Jewish joke about the drunk looking for his keys where the light is better rather than where he has lost it.

It is for these reasons that, in direct contradiction to the notion of “pure” indices of effect size independent of the raw score, it is argued that simple raw scores make possible a less ambiguous statement of effect size. For example, a procedure that increases the linguistic intelligence of children might report the change in raw points on the Similarities subscale of the WISC-III. A treatment for individuals suffering somatization disorder might report the change in raw Hypochondriasis scores on the MMPI-2. And a program for treating children with severe emotional and behavioural problems might report change in raw Total Problem Score on the CBCL. A distinct advantage of raw scores is that they are free from any misleading claim of their “pure” nature which deflect attention from the scientific, social, instrumental, methodological and substantive context in which the results arise.
It has been argued in this section that the current methods which measure “clinical significance” of psychological test data conflate two separate issues: a quantitative aspect of outcome and its “clinical significance”. The quantitative aspect involves either of two logically distinct notions of outcome: the change from pre to post treatment (i.e. effect size), and the post-test result (i.e. level of post-treatment pathology). The importance of maintaining this distinction will be discussed further in the next section [2.5]. It has been argued that the “clinical significance” of either of these two types of outcome, however, is a qualitative issue. Clinical significance is a broad term, and arises from the substantive context in which the results occur, and this has been illustrated with several examples.

It has been argued that for these reasons no one procedure or index, or set of procedures and indices, can be definitive of “clinical significance”, and that the appropriation of the term “clinical significance” in any such way is misleading and reductive. If adopted prescriptively, such a definition limits the study of clinical phenomena. Furthermore, some examples have been provided from the literature of the uncritical way in which Jacobson’s criteria of “clinical significance” has already been applied and reported. This raises concerns about misleading and self-serving results purporting objectivity on the basis of its ostensibly “mathematical” procedures.

Finally, the proposal for a “pure” index of effect size to facilitate comparisons between studies is examined. It has been argued that the most straightforward way to present quantitative
outcomes – the raw test score most readily lends itself to the unavoidable qualitative consideration of its clinical significance, and is free of misleading claims of its “pure” nature.

(In the present study, described in Chapters 4, 5 and 6, no formulaic definition of “clinical significance” is adopted, and the discussion of outcomes involves raw test scores).
2.5 Failure to maintain distinctions between different kinds of outcomes

We come now to the final set of conceptual and methodological issues, about which there are confusions seriously impeding progress, to be discussed in this chapter. It is the uncritical way in which different types of outcomes are often run together in the literature. This is examined under two categories: failure to maintain distinctions between logically different meanings of outcome, and failure to maintain distinctions between outcomes arising from different sources of information.

2.5.1 Two logically distinct meanings of the notion of “outcome”

Here, the discussion will centre on confusion between the two basic, logically distinct kinds of outcome derived from psychological measurement instruments, change in pathology and the level of post-treatment pathology, as outlined in the previous section [2.41].

A large enough change in the patient’s symptoms or condition or complaints over the course of treatment is called a good outcome. At the same time, the achievement of a particular sort of result when the symptoms or condition or complaints are within the “normal” or “non-clinical” range, which might be referred to as a “cure” or “remission” in the popular view of medicine, is also called a good outcome. Yet the logical difference between the two kinds of outcome is obvious. While it has been pointed out in the previous section that the adoption of Jacobson’s dual criteria of clinical significance as a general definition of clinical significance is unnecessarily restrictive and obscures the complexity of the issue, Jacobson at least is very clear that two different criteria are being discussed. This is not always so in the general discussion of research results which appear in the literature, where outcomes of studies are discussed without
reference to which general kind of outcome is being reported. This might be best illustrated with an example.

In *What Works for Whom: a Critical Review of Treatments for Children and Adolescents*, Fonagy, Target, Cottrell, Phillips & Kurtz (2002) only nine studies pertaining to inpatient and day patient (“partial hospitalization”) treatment – having met their criteria for relevance, outcome and design – are reviewed. Of these, only one study referred to the outcomes of a comprehensive treatment with a heterogenous sample of troubled youths. I refer here to treatment characterised by simultaneous intervention at individual, family and school levels, with youths showing various emotional and behavioural problems, rather than, for example, a sample of males showing anti-social behaviour. That is, types of treatment similar in important ways to that offered at Arndell Unit during the 1990s. Interestingly, this study (Rey, Denshire, Wever & Apollonov, 1998) took place at the day program at Rivendell Unit, New South Wales, Australia, a “sister” program to Arndell.

While meeting the authors’ criteria for inclusion in the review, there are quite significant weaknesses in the study. These need to be considered before moving to the salient point which pertains to confusions between two logically distinct meanings of the notion of “outcome”, and the way in which this confusion is compounded in the rereporting in Fonagy, Target, Cottrell, Phillips & Kurtz (2002) of the original report by Rey, Denshire, Wever & Apollonov (1998).

First, there are problems with the methods utilised in this study for determining outcomes. Adolescents were designated as showing a poor outcome in this study “if they had committed suicide, had been in prison, were expelled or dropped out of school, were on welfare benefits, were living with the sexual partner, were pregnant or had a child, were living in a refuge or
homeless, had been admitted to hospital for mental health treatment, abused alcohol or drugs, or were unemployed” (Rey, Denshrie, Wever & Apollonov, 1998, pp 45 – 46). The heterogeneity of these items, the implicit moralism, and insensitivity to the developmental and familial contexts of the individual case limits the meaningfulness of this finding. For example, given the age range at followup of 13 to 19, “living with a sexual partner” has a very different clinical significance for, say, a delinquent 13 year old from a chaotic family and a 19 year old who previously suffered separation anxiety in his enmeshed family. Similarly, “living in a refuge or homeless” would have a different clinical significance depending on the family conditions the youth left behind.

Second, there are problems with the methods utilised in this study for assessing the reduction in CBCL scores. The general conclusion that “there was a considerable reduction in CBCL scores between assessment and follow up” is based on the “size” of the decrease which was “significant (p<0.05)” on all but two subscales, with a “particularly large” reduction on three subscales with \( p \) values < 0.001. The common confusion between statistical significance and effect size is evident.

Third, there are problems with the research design. The control group was constructed from adolescents referred to Rivendell Unit but who were “not admitted to the day program because they refused, lived too far away, no vacancies were available, or for a variety of reasons were not considered appropriate for that specific program” (p. 43) who then went on to “a variety of other treatments” (p. 46). Given the heterogeneity of this group, and that an unspecified number may have gone on to treatment at least as intensive as Rivendell at other specialist units operating in Sydney at that time with quite different treatment philosophies, approaches and
internal structure (Arndell in North Ryde, and Redbank in Westmead), it is unclear what was being controlled by this group.

Fourth, there are problems with there being insufficient data made available for the reader to come to their own view about the clinical significance of the change in symptoms which took place. For example, the information provided does not include pretest and posttest means on CBCL scores which might assist the readers to come to their own view about the size of the change achieved in problem behaviours. Furthermore, the social situation of the adolescents pre-treatment, which would provide some context for the “poor” social outcomes reported at follow-up, is not provided. It is also noted that while the section of the paper referring to decrease in symptom scores is headed “Change in CBCL and YSR” scores, there appears to be no mention of the latter.

These weaknesses show the difficulty of setting inclusion criteria for the empirical review of a large area, and more generally, the problem of reducing the overwhelming number of published reports to a manageable size. However, as stated, the salient point pertains to confusions about the reported outcome of the study.

The paragraph in Fonagy et al (2002) summarising this study is reproduced in full.

“Rey et al. evaluated an Australian multimodal partial hospitalization program at 3-year follow-up (Rey, Denshire, Wever & Apollonov, 1998). Parents and children were interviewed separately. Those adolescents who attended the programme (n = 38), many with diagnoses of CD and CD/ADHD, functioned better overall and were more satisfied with the treatment than the controls (n = 35). The groups were
matched on delinquency and aggressivity scores as well as age and gender. However, the study yielded **poor outcome for both the treated and treatment-as-usual control groups**. Initial diagnosis of CD and high delinquency and total problem scores predicted poor outcome” (p. 179, emphasis added).

The abstract from Rey, Denshire, Wever & Apollonov (1998) is as follows.

“Adolescents (N = 38) who attended a multi modal day program were matched with a control group (N = 35) for age, gender, year of assessment, and on delinquency and aggressive scores. Parent and child were separately interviewed three years after initial assessment. Adolescents who attended the program functioned better overall and were more satisfied with treatment than controls. **Both groups showed a poor outcome and a decrease in symptom scores**. A diagnosis of conduct disorder, high delinquency scores and high total problem scores predicted a poorer outcome” (p. 42, emphasis added).

While the two passages are very similar, the summary statement with regard to outcomes (in bold) is quite different. Fonagy et al do not qualify whether “poor outcome” refers to the level of *change in pathology* or the level of *post-treatment pathology* found at followup. Rey et al, in fact, reported “a poor outcome and a decrease in symptom scores”; the poor outcome refers to the proportion of adolescents deemed to be in unsatisfactory social situations at followup, and the decrease in symptom scores refers to differences between the CBCL administered at assessment and at followup. It can seen that the finding reported by Rey et al of **“a poor outcome and a decrease in symptom scores”** - problems still evident post-treatment despite
improvement in symptoms – becomes the much less optimistic, unqualified conclusion in Fonagy et al of “poor outcome”.

The clinical and pragmatic significance of the difference between these two conclusions is not trivial, especially in the context of social accountability, allocation of resources and cost-effectiveness which the survey by Fonagy et al aims to address. The conclusion by Rey et al might lead, for example, to interest about the factors associated with the improvements achieved in adolescents and how they might be further increased, and to practical considerations about the increase in costs of an enhanced treatment program versus the long-term financial and social costs of adolescents with poor adjustment at age 13 to 19. The summary by Fonagy et al more readily lends itself to therapeutic nihilism, where, for example, the considerations of reducing the costs of treatment may become prominent; health organisations and government departments are, understandably, unlikely to show interest in expensive day programs with “poor outcome”.

Furthermore, statements about predictors of outcome have quite different implications depending on which of the two types of outcome is at issue. The conclusion by Rey et al that “a diagnosis of conduct disorder, high delinquency scores and high total problem scores predicted a poorer outcome” (from the abstract, page 42) refers to their unremarkable finding of an association between pre-treatment pathology and post-treatment pathology. However, when “outcome” is left ambiguous, as in the excerpt from Fonagy et al above, the reader may mistakenly assume that “poor outcome” refers to little or no change in pathology. The original unremarkable finding can then be misunderstood as indicating that certain adolescents

---

12 While it is reported that an association was found between particular types of psychopathology (conduct disorder and delinquency) and post-treatment pathology, it is difficult to interpret this result as it appears that the type of pathology may have been confounded with the severity of pre-treatment pathology. That is, it may simply be the case that the adolescents diagnosed as conduct disorder were the most severely disturbed.
(identified by their externalizing behaviours and/or their general level of disturbance) are not responsive to treatment (i.e. do not show change), again leading to therapeutic nihilism.

This type of carelessness in maintaining distinctions between the two kinds of outcome is not rare, and it is not likely to be trivial in their effect. These errors can be found in the work of influential researchers, misinforming quite specific policy recommendations. To take another example, Kazdin and Weisz (1998) addressing child and adolescent treatments in the special section on “empirically supported psychological therapies” in the February 1998 issue of the *Journal of Consulting and Clinical Psychology*, make the following statement about predictors of outcome for “problem-solving skills training (PSST)”, which they regard as “an extremely promising approach” (p. 25) for children showing externalizing behaviour.

> “Some evidence suggests that... youths who have comorbid diagnoses, academic delays and dysfunction, and lower reading achievement and who come from families with high levels of impairment (parent psychopathology, stress, and family dysfunction) **respond less well to treatment** than youths with less dysfunction in these domains (Kazdin, 1995; Kazdin & Crowley, 1997)” (p. 25, emphasis added).

In fact, Kazdin & Crowley (1997) tested the strength of association between these predictors and post-treatment scores on the CBCL and the TRF. Furthermore, the “total number of symptoms” was “used as an index of comorbidity” (p. 203). While it is difficult to understand “respond less well to treatment” as anything other than a reference to lack of change in these youths, the study can only be said to have, again, made the unremarkable finding that those showing the most difficulties pre-treatment showed the most difficulties post-treatment. Given the confounding between, on one hand, comorbidity and, on the other hand, severity of pathology, nothing
meaningful about the relationship between comorbidity and outcome can be concluded from this analysis.

The authors, however, draw the following conclusions from their findings:

“In relation to clinical care the success of treatment will not only depend on identifying more effective interventions, but also on directing cases to available treatments from which they are likely to profit” (p. 205).

It is discomforting to ponder whether the practical implications of their mistaken conclusion may be to cause the most disturbed children, from the most disadvantaged family situations, to be directed away from treatment. While the inclusion of pre-treatment pathology among predictors of post-treatment pathology in multiple regression analyses readily leads to the unremarkable finding that “the best predictor of post-treatment functioning is pre-treatment functioning” - a pithy conclusion easily remembered, if not well understood, by clinicians - change in the level of pathology from pre to post-treatment is an entirely separate issue. It needs to be remembered, for example, that a uniformly powerful treatment effect will also maintain the correlation between pre and post-treatment morbidity!

2.52 Different sources of information about outcome

There are also confusions between the different kinds of outcome arising from different sources. The focus here will not be on the consolidation of data arising from different methods or different instruments, but on differences due to different observers of the same outcome variable. While the following discussion takes the Achenbach scales as an example, it has been
reported that there has been a general trend in the field of psychotherapy research towards the gathering of multiple sources of information about outcome (Lambert & Hill, 1994, p. 74 - 75).

Yet, “The fact that differences in outcomes have been found to be a function of source rather than content has been replicated across a variety of scales and patient populations and across three of four decades suggests that this finding is very robust”, and “few studies recognize or deal adequately with the complexities that result from divergence between sources..” (Lambert & Hill, 1994. p. 80 - 81). The development of parallel forms of the Achenbach scales for use by different raters, and their wide adoption in multiple-perspective research, means that there is quite a sizable literature available which is particularly well suited to a discussion of differences in outcome associated with source.

The Achenbach scales in their parent (Child Behavior Checklist, or CBCL), teacher (Teacher's Report Form, or TRF) and self-report (Youth Self-Report, or YSR) forms are probably the most well known and widely used instruments for rating the behaviour of children/adolescents with emotional and behavioural problems in clinical settings, in the research literature and in epidemiological surveys (Achenbach, 1991). Extensive norms are available, including comparisons of data from a broad range of cultures across the world (Crinjen, Achenbach & Verhulst, 1997), and its broad acceptance as a standard has generated considerable discussion and research of its applications and psychometric properties. However, the notion that any of the Achenbach scales possibly represents a “gold standard” in any application is now generally considered unsophisticated, and there is much discussion with regard to the circumstances under which, for example, the rating by a mother might be preferred to one by her child, or a rating by a teacher might be preferred over that by a mother (Bird, Gould, Staghezza, 1992). There is also a growing understanding that disagreements between raters are not random errors (Youngstrom, Loeber & Stouthamer-Loeber, 2000).
Three examples are provided to illustrate the wide range of issues which arise with regard to differences in outcome which are “a function of source rather than content” on the Achenbach scales.

- It appears that youths in the community report more externalizing problems than their parents, and the reverse is found among clinic-referred youths (Sawyer, Baghurst & Mathias, 1992). The authors suggest that this pattern may be due to clinic-referred youths showing more readily observable problem behaviours, or the clinic-referred youths being less likely to report problem behaviours because they had been punished for these behaviours in the past, or the referral process having “influenced the way family members completed their checklists” (p. 448). The way in which the latter might take place is left unspecified.

- It has been found that depressed mothers report more problems relative to their children’s teachers than non-depressed mothers (Chilcoat & Breslau, 1997). It has been suggested that this may be due to depressed mothers overstatement their child’s problem behaviours, or that these children may show more problem behaviours at home than at school (Biederman, Mick & Faraone, 1998; Ingersoll & Eist, 1998; Chilcoat & Breslau, 1998).

- A replication in Sydney with a sample of 1300 children (Hensley, 1988) of the original US normative study of the CBCL with a sample of 1300 children (Achenbach & Edelbrock, 1981) showed very much higher scores (total problem scores of 31.6 versus 20.1). This has not been able to be explained by the differences between the two
samples with regard to socio-economic status, urbanization, or proportion of immigrants (Achenbach, Hensley, Phares & Grayson, 1990). The authors seem unsure what to make of this, even suggesting at one point that “the difference of 11.5 points in mean total score represents a small proportion of the total range of 240 points on the problem scale”, and alternating between the view that Sydney children showed “high rates of problems”, that “Sydney parents generally report more problems for their children”, and that there might be “consistent differences in adult’s perceptions of child problems between the two countries” (p. 284, italics added). Interestingly, the Western Australian Child Health Survey (Zubrick, Silburn, Garton, Burton, Dalby, Carlton, Shepherd & Lawrence, 1995) of 1372 children relied largely on the CBCL, TRF and YSR scored according to US norms for its estimate of mental health morbidity. Its data was made available for a comparison of 12 cultures (Crijnen, Achenbach & Verhulst, 1997) where, in sharp contrast to the Sydney data, the total problem score was consistently lower by a small margin than that obtained from US parents over the age ranges of six through eleven, and twelve through seventeen (3.4 points and 2 points, respectively).

A sophisticated view of the data obtained from the administration of the Achenbach instruments would hold that they represent the product of a complex procedure that involves a communication from the rater regarding their perception of the subject's behaviour. Such a process is vulnerable to distortions, biases and demand characteristics operating on the rater at either point. That is, inter-rater disagreement may reflect actual differences between two individuals in their perception of the child’s behaviour (or in the case of the YSR, the child’s self-perception), or differences in what they wish to communicate about the behaviour (and therefore their child and themselves), to the individual or agency requesting this information.
While some of these factors are readily acknowledged by Achenbach, for example, in the section of the CBCL manual on “practical applications” - where parental disagreement is treated as a clinically significant anomaly to be explored further (Achenbach, 1991, p. 167 -168) – they are largely absent from his discussion of the “research use” of the instruments in the CBCL manual, or the Integrative Guide for the CBCL, YSR and TRF (Achenbach, 1991). It is also absent from the YSR manual where the reported validation of the instrument through administration to referred and non-referred samples shows the YSR to have less consistent results in the expected direction than the CBCL or TRF.

The issues raised by these three examples will be discussed in detail.

With regard to the findings by Sawyer, Baghurst & Mathias (1992), it is not surprising that youths provide ratings less consistent with their clinical status (ie. referred vs non-referred) than that of their parents who determine their child’s clinical status. With regard to what adolescents may wish to communicate it might be expected, for example, that compared to non-referred adolescents clinic-referred adolescents will be less willing to comply with the requirements of the assessment task, or less prepared to be frank about their problems and areas of poor competences with clinicians in positions of authority over them. With regard to perception, there are likely to be real differences between the self-perceptions of these adolescents and the perceptions of their parents. It is expected in families with troubled youths that there is an intense battle with regard to the “true” location of “the problem” (i.e. who in the family is “sick” and needs treatment, who in the family is “at fault”, and so forth) and the family’s completion of Achenbach instruments may well reflect this.
With regard to the findings by Chilcoat & Breslau (1997), where ratings by mothers and teachers disagree, it is not possible to determine on the basis of the CBCL and the TRF to what degree this might, for example, reflect a *communication* by the mother of a troubled child-parent relationship (i.e. a “cry for help”), actual differences between the child’s home and school behaviours, the *perception* of the mother stuck in this troubled relationship, or some combination of the above.

With regard to the findings by Hensley (1988), a possible explanation of the discrepantly high CBCL scores obtained from parents in Sydney compared to parents in the US normative study (as well as parents in Western Australia) may lie in what appears to have been an important variation in data collection procedure in the Sydney study.

“Within each targeted suburb, streets were chosen randomly from street directories. The procedure was a simple doorknock….. To increase the efficiency of the yield in each street and to minimise bias against working mothers, we asked cooperative residents to nominate houses in the vicinity of their own in which there were resident children. These informants were not necessarily parents, nor were they necessarily known to subsequent subjects, and each selected street was independent – that is, snowballing was avoided as much as possible. We returned to empty nominated houses and to others where there was other evidence of child occupancy (swings, bicycles, etc) on weekends and early evenings” (Hensley, 1988, p. 373).
It might be expected that in comparison with “cold-calling”, which appears to have been the method in the US study, referrals by neighbours may have more warmly disposed the families to the interviewers. This combined with another variation from US procedures, the “appeal to nationalism” where “To survey-weary questions about the usefulness of such a study we replied that much information about children in this country is from overseas and that we need specifically Australian information for proper understanding and planning for our children” (Hensley, 1988, p. 373 – 374). It is also noted that there was “a remarkably enthusiastic research assistant” (p. 373). It might then be expected that the Sydney procedure may have “induced parents to ‘pathologise’ their children in the hope for better child-oriented services” (p. 378) – a possibility which the author mentions and rejects.

The basis for this rejection is that the Dutch studies used a similar appeal while remaining very comparable with the US results. With the publication of the quite large differences between 12 different cultures across the world (Crijnen, Achenbach & Verhulst, 1997), and the large differences between the Sydney and Western Australia results, a broader approach to the consideration of differences between studies appears to be called for. It might be suggested, for example, that as well as the possible demand characteristics of the approach taken by Sydney investigators towards increased “pathologise”–ing by parents, a “remarkably enthusiastic” approach by investigators calling on individuals on the recommendation of their neighbours may have lead the parents to take a more relaxed, less defensive, more frank attitude to communicating the behaviour of their children. It can be seen, in this respect, that the item on which the difference between Sydney and US endorsements was most prominent was “thinks

---

13 The original paper (Hensley, 1988) does not include the nomination of houses where children resided among the “exceptions to the American procedure” (p. 373), while the “descriptions of samples” in the later paper (Achenbach, Hensley, Phares & Grayson, 1990, p. 268 - 269) suggest that it was a procedural deviation.
about sex too much”, which was reported by 39% of Sydney parents compared to only 2% of US parents (Achenbach, Hensley, Phares & Grayson, 1990, p. 270). It is also the case that Sydney parents voluntarily added more problems than US parents on items 56h (“other physical symptoms”) and 113 (“other problems”). These sorts of differences in responsiveness suggest that rather than the differences in, for example, sex-related items representing “an area of greater concern for Sydney than American parents” (p. 284) as suggested by the authors, it may represent the opposite – Sydney parents may have been more likely to endorse these items because they were less concerned, and more willing to communicate to the investigators about their children’s behaviours.

Given the issues arising from these three examples, illustrating some of various ways in which differences in outcomes on Achenbach ratings instruments may reflect important, non-random, and quite large effects of the interaction between respondent and test administrator, it is of concern that there are still very crude methods for “combining” the data from different sources.

For example, what has been put forward as the most promising method for “aggregating” Achenbach scale data from parents and teachers (Bird, Gould, Rubio-Stipec, Staaghezza & Canino, 1991; Bird, Gould & Staghezza, 1992) has been aimed at the development of “the most sensitive screen for ‘caseness’”. It appears, however, that ‘caseness’ is determined by psychiatrists mainly on information provided by the parents rather than their own observations of the child over time. The concerns with this particular approach are as follows.

14 As it is reported that “a comparison of inter-interviewer and inter-informant (mother, father, or other) total scores yields no significant differences” (p. 378) the effect does not appear to be attributable to the enthusiasm of any particular interviewer.
First of all, this involves a collapse of an empirically-derived, dimensional taxonomy based on factor analytic studies of the psychopathological symptoms of childhood (Achenbach, 1985; Quay, 1986) into the conventionalistic, discrete groupings of psychiatric diagnosis, sweeping aside the contrasting definitions of abnormality in the two approaches. In the former taxonomy, children scoring beyond arbitrary statistical cut-offs on the Total Behavior Problem Score – those at the 82\textsuperscript{nd} percentile (T = 60) are considered to be in the “borderline clinical” range and those at the 90\textsuperscript{th} percentile (T = 63) are considered to be in the “clinical” range (Achenbach, 1991) - are defined as showing behaviours significantly deviating from the mean. In the latter taxonomy, a child is categorised into one or more distinct mental disorders according to the definitions in the current Diagnostic and Statistical Manual of Mental Disorders published by the American Psychiatric Association. While each diagnostic system is arbitrary, each is internally consistent and can, at least in principle, be applied in a reliable and consistent manner. However, collapsing data gathered by one system into the other leads to quite meaningless results under the guise of objectivity. For example, the Western Australian Child Health Survey (Zubrick, Silburn, Garton, Burton, Dalby, Carlton, Shepherd & Lawrence, 1995) scored the CBCL and TRF using US norms and defined “mental health morbidity” in the cases that showed T scores of 60 or more (i.e. the 82\textsuperscript{nd} percentile) on either test. The survey then reports that 20% of male children and 15.4% of female children in Western Australia have “mental health problems” (p. 35). They do not mention that by using local norms and selecting a more rigorous definition of, say, T = 63 (90\textsuperscript{th} percentile) on the CBCL they can confidently estimate that 10% of the child population show mental health morbidity (without conducting a survey!)

Secondly, aggregating ratings from parents and teachers – as also took place in the Western Australian Child Health Survey - means that information about potentially important dimensions of child problems pertaining to the setting of the problem behaviours and therefore the
pervasiveness of problem behaviours, is lost. While the CBCL and the TRF are indicators of the severity of problem behaviours reported by parents and by teachers, there are likely to be important differences between the child who shows severe problem behaviours at home, the child who shows severe problem behaviours at school, and the child who shows severe problem behaviours in both settings. This remains a neglected area of research.

Two general conclusions follow from the foregoing considerations about the failure to maintain important distinctions between different sources of outcomes.

First, while interest in “projective” tests of personality have receded over the last 100 years and replaced by an interest in “objective” tests of personality and behaviour, the ways in which the latter methods unavoidably involve a complex process of communication between the test administrator and testee regarding the testee’s subjective perceptions has been largely neglected. That is, the “projective” aspects of ostensibly objective instruments remain ignored. For example, contemporary developments in the Rorschach inkblot method reflect a sophisticated understanding of the response process based on specific research on the following: the range of perceived responses; the considerations of social desirability and perceptual accuracy (i.e. the goodness of fit between the response and the inkblot) which inform the articulated response; and the degree and quality of articulation of the response (Exner, 1978, 2002). To provide another example from projective testing, the present recommended procedure for administration of the Rorschach method – where the examiner and subject sit side by side - developed from studies which showed that this reduced the examiner influence on testees evident in face-to-face seating (Exner, 1986, 2002). It needs to be emphasised that this influence was deemed to occur through non-verbal cues. The results of this area of research are quite remarkable. It has been shown that naïve graduate students trained to administer the Rorschach, divided into two groups sets and
receiving different instructions – either that experienced examiners always elicited more human than animal responses, or that experienced examiners always elicited more animal than human responses – produced responses from naive examinees in the expected direction. It is reported that there was no evidence that verbal conditioning was taking place (Masling, 1965). Exner (1986, p. 64) reports further modifications on this design which showed that “postural, gestural, and facial cues would occur when the subjects [testee] gave a response related to the [mental] set [of the examiner]” and that the effects of the different instruction to the examiners disappeared when side-by-side seating between examiner and examinee was introduced. In contrast, the CBCL normative studies in the US and Sydney both took place with the interviewer reading aloud each item and writing down their response.

(The steps taken to maintain the same rater at different rating points, and to protect the rater from the demands of the institution, is a feature of the present study, described in Chapters 4, 5 and 6).

Second, while it is often stated that in psychotherapy outcome research it is necessary to have multiple perspectives, especially when children are in treatment (Kazdin, 1994, p. 574), there remains the problem of what is to be done when the perspectives are in disagreement. More specifically, current methods of aggregating reports by parents and teachers leave neglected the question of whether clinically important differences exist between youths of whom severe problem behaviours are reported by parents at home, those of whom severe problem behaviours are reported by teachers at school, and those of whom severe problem behaviours are reported in both settings. It is argued that a more defensible alternative is to treat the differences as potentially meaningful, and worthy of study. (This is the approach adopted in the present study, described in Chapters 4, 5 and 6).
CHAPTER 3.
LITERATURE REVIEW: SUBSTANTIVE FINDINGS ABOUT INTENSIVE, PSYCHO-SOCIAL, DAY & RESIDENTIAL TREATMENT FOR SEVERELY EMOTIONALLY DISTURBED CHILDREN

3.1 Introduction to this Chapter

This chapter is concerned with substantive findings about intensive, psycho-social, day & residential treatment for severely emotionally disturbed children. Programmes conducted primarily on an outpatient basis, at home (e.g. Bidder, Gray & Pates, 1981), or within regular schools (e.g. Rickel, Smith & Sharp, 1979) will not be included in this review. Programmes catering predominantly to the intellectually disabled, or neurologically damaged will not be included in this review. Nor will programs which are entirely or predominantly custodial (i.e. within the juvenile justice agencies) or educational. Similarly, mental health facilities for children and adolescents offering only residential placement and relying entirely or predominantly on psycho-pharmacological management of problem behaviours will not be included. Centres which offer short-term emergency accommodation will not be included. While it may be unhelpful to draw distinctions between youths who are channelled to a justice facility, a mental health facility, or a special education program, the focus of the present review is on intensive psycho-social treatment programs.

It is organised in the following way. In the next section, empirical reports pertaining to the effectiveness of day treatment programs are examined [3.2] and in the following section, empirical reports pertaining to the effectiveness of residential treatment programs [3.3]. Next, reports pertaining to predictors of outcome are examined [3.4]. The final sections of this chapter contain a summary of substantive findings and methodological problems [3.5] and the indications for productive lines of future inquiry [3.6].
There are a considerably larger number of studies which address effectiveness with respect to day treatment than with residential treatment, and one study was found which compared the effectiveness of day versus residential treatment within the same setting. That study – by Goldfarb, Goldfarb & Pollack (1966) - will be discussed under studies pertaining to predictors of outcome [3.4].
3.2 Studies pertaining to the effectiveness of day programs.

The studies in this section address with regard to day programs the question: how effective is the treatment under examination?

This section is organized according to the age of the children and adolescents concerned. Most of the studies to be considered here involve institutions catering to a particular age group, with a few institutions catering to a very wide age range. For the purposes of clarity, the population of each study will be described in terms of the school groups to which they roughly correspond: pre-school (3 years to 6 years), infant (6 years to 8 years), primary (8 years to 12 years), junior high (12 years to 15 years) and senior high (15 years to 18 years).

As for the types of emotional and behavioural problems shown by the children and adolescents, or the psychiatric diagnostic categories to which they are allocated, this varies considerably among the institutions included in this review, with most catering to a rather mixed group. For the purposes of uniformity, the problem type of the children and adolescents will be described mostly with reference to the following rough categories\textsuperscript{15}: internalizing, externalizing, psychotic and organic.

The general format in this section is as follows: a brief description of the treatment program; a brief description of the study; a summary of the authors’ interpretation of their results; a critical

\textsuperscript{15}The terms internalizing and externalizing represent a distinction between children and adolescents with anxious, depressive, withdrawn or inhibited presentations, on one hand, and those showing aggressive or antisocial behaviour, on the other. Alternatively designated acting-in/acting-out, or neurotic/conduct disordered, this distinction is widely considered to be clinically meaningful. Groupings of this sort have been identified in factor analytic studies of the psychopathological symptoms of childhood (Achenbach, 1985; Quay, 1986). As mentioned [2.52], this approach represents an alternative, empirically-derived, dimensional taxonomy to that offered by the conventional, discrete groupings of psychiatric diagnosis.
comment on what confidence we can have in this interpretation; and comments, where appropriate, on any substantive conclusions which might be drawn from reported results.

This section will commence with programs for pre-schoolers [3.21], then programs for infant school aged children [3.22] and then programs for infant and primary school aged children [3.23]. It will conclude with programs catering for a broad range of age groups [3.24].

The names of institutions are bolded.

3.21 Programs for pre-schoolers


Eisenberg, Lansdowne, Wilner & Imber (1962) reported on the effectiveness of the Children's Guild, Baltimore, USA, a therapeutic nursery (ie. pre-school) programme which provided "emotionally disturbed children a group experience under the close supervision of qualified teachers and skilled psychiatric case work to their parents" (p.18). The majority of children showed internalizing and externalizing problems, with another 20% described as organic and 5% described as psychotic. Median attendance was for 6 months.

First of all, a comparison was made between a class of 19 children (proportion of boys to girls not reported) admitted for treatment (to be called here treatment group A) and a normal control
of 218 children attending regular pre-schools (control group A) on two rating scales designed for this study: a symptom check list and a "health inventory". The ratings were completed by the child's teacher. As expected, the treatment group was found to show more disturbance and less competence than the control group.

Secondly, ratings were made on 16 children (10 males/6 females) who had completed treatment in previous years and were now attending public schools (treatment group B) and compared to ratings on a control group of 136 children from the same schools, matched for sex (control group B). Among the boys, significant differences were found between treatment group B (n=10) and control group B (n=72) on the symptom check-list and differences "at the border of significance... (p<0.10>0.05)" (ibid, p.23) were found on the health inventory. Among the girls, no significant differences were found between treatment group B (n=6) and control group B (n=66) on either scale.

The authors conclude that these results are "presumably a reflection of a greater degree of recovery in the girls than the boys, since both boys and girls had differed from the co-op controls in the class studied at admission" (p.23). Furthermore, the authors suggest that it provides "preliminary evidence of clinical improvement in the status of the patients after the guild experience, though the lack of suitable controls precludes ascribing the change to the therapeutic program alone" (p.27).

There are problems with the design of the study, the data analysis and the interpretation of the results. This interpretation of the results assumes equivalence between the treatment groups A and B and also between control groups A and B. Leaving this aside, the main objection to these

16 The inventory appears to contain items pertaining to what would now be referred to as "social
conclusions lies with the common misinterpretation of results of statistical tests of the difference between two means – as representing the size of differences between groups (discussed in 2.41). First, an unsound argument is made for the equivalence of groups on the basis that no statistically significant differences were found. It is unclear, for example, whether the finding of a statistically significant difference between treatment and control groups for boys but not for girls is attributable to the difference in statistical power due to the smaller number of girls. Second, a similar problem exists for interpreting the finding of a statistically significant difference between treatment group A and control group A, but not between treatment group B and control group B, as N was considerably larger for the A groups.

The authors' conclusion that the results represent "preliminary evidence of clinical improvement" is not supportable by their analysis of the data. Furthermore, their qualification that "the lack of suitable controls precludes ascribing the change to the therapeutic program alone" is misleading in its implication that change has in fact been demonstrated for the treatment group. With not even average group scores being reported, it is not possible to assess the size of differences between groups.

Mitchell, Rothwell & Burtenshaw (1975) reported on the effectiveness of the **London Hospital Pre-School Unit**, London, UK, which provided treatment four days a week for an average of six months to children showing mainly internalizing and externalizing problems. The focus of treatment was on "the dynamics of the interaction between mother and child" (p.390). The mothers and children were together in the mornings and the mothers attended a group in the afternoons, with up to seven families attending at any one time.

---

competence" or "social skills".
The study involved a comparison between 14 children who received treatment and 7 children whose family declined the offer of a place. The proportion of males to females not reported. There were three measures of outcome. Mothers were asked at follow-up to retrospectively rate the severity of their child's problem behaviours at the time of initial referral to the Unit and then asked to rate their child’s current behaviour, with the total score at follow-up subtracted from the total score at registration to produce a change score. (The rating scale used is not identified - presumably it was specially constructed for this study - nor is there an indication of how long after discharge follow-up took place). In addition to this, the child's class teacher at follow-up (only) completed the Rutter Rating Scale of Behaviour Disorders (Rutter, 1967), and the Social Behaviour Rating Scale (Herbert, unpublished) which provides four scores (Behaviour Disorder, School Competence, Dependency and Peer Relations).

Data analysis consisted of statistical tests of the difference in mean scores between the treatment and non-treatment groups, with the following results. While the mothers' change scores were significantly higher for the treatment group, the teachers' ratings at follow-up showed a significant difference between treated and untreated groups on the Dependency score of the Social Behaviour Rating Scale with the treatment group showing greater dependency, which is contrary to expectation. The authors concluded that "treated children showed a greater reduction in reported problem behaviours at home than did control children" but that "attendance at the unit had a limited effect on behaviour at school" (p.394).

This conclusion of a limited but nonetheless positive effect of treatment might be challenged on several grounds. First of all, as with all designs involving a non-equivalent control group (ie. quasi-experimental design), a threat to internal validity lies in pre-existing between-group differences which might account for the difference in outcome. In the present case, it might be
expected that families of children referred for treatment who turn down the offer of a place would differ from those accepting treatment in ways which are prognostically relevant. For example, the mothers may be less motivated for treatment or more defensive about problems they might be experiencing with their children, or the fathers may be less supportive of participation in the programme, and so forth. Factors of this kind lie well within the Unit’s stated clinical orientation. Secondly, retrospective ratings by patients can be expected to be particularly vulnerable to bias due to the demand characteristics of the situation (Rosenthal, 1976). No attempt appears to have been made to address any of the aspects of the potential influence of researchers' expectations on mothers' ratings, such as (a) the mothers' perceptions of an association between the research interviewers and the treatment Unit, (b) whether interviewers were blind to the researchers' expectations with regard to results, or (c) whether the interviewers were blind as to which group the participants belonged.

These issues present serious difficulties for the authors' interpretation of the result from statistical tests on maternal ratings as indicating a positive treatment effect. Interestingly, they do not interpret the higher Dependency scores in average teacher ratings for the treatment group as indicating a negative treatment effect. This is an unexpected and interesting result, but given the non-equivalence between groups and the absence of pre-treatment ratings by teachers it is difficult to interpret. These difficulties notwithstanding, it would be of interest to know about the size of the change and the post-treatment functioning of the children relative to norms. The

---

17The authors do attempt to demonstrate equivalence between the two groups with regard to (a) mothers' base-line ratings of problem behaviours, and (b) what the authors refer to as "stress factors". For the latter, clinical notes were retrospectively rated on 17 items (on a scale presumably specially constructed for this study) which were aggregated to form four scores: "stresses in family structure", "health", "housing" and "work and finances". Aside from the faulty logic, as discussed, of arguing that the groups are equivalent on the basis that no statistically significant differences were found, these procedures do little to address concerns about group differences associated with family motivation for treatment, and so forth.
use of a self-made rating scale without established psychometric properties or norms provides is unhelpful in these respects.

Woollacott, Graham & Stevenson (1978) reported on the therapeutic effectiveness of the psychiatric day centre at the Hospital for Sick Children, London, UK. In addition to internalizing and externalizing problems, some of the children showed developmental delay and about 10% are described as "autistic or show marked autistic traits" (p.350). The description of the treatment focus and the approach to treatment here was similar to that of the London Hospital Pre-School Unit: "to improve the level of disturbance mainly by helping parents to become more aware of the emotional needs of their children, improving communication between parent and child, and encouraging parents to obtain a more accurate view of the child's level of development... most emphasis was placed not on the specific removal of symptoms but on the provision of a generally caring and supportive environment with understanding staff who could act as models with whom parents could identify" (p.350). Children attended with their mothers and younger siblings, and sometimes with fathers as well. This setting also provided parent groups, and it is reported that most parents attended individual sessions with the social worker. However, attendance was usually only one day a week, with 8 to 10 children attending each day, so that 40 to 50 children were seen each week.

The study involved a comparison between a group of 25 children (72% male) who received treatment and two groups of matched controls selected from a community sample available from a separate epidemiological study. The first control group (48% male) was selected by matching on a measure of severity of disturbance (score on the Behaviour Screening Questionnaire [Richman & Graham, 1971] administered to mothers by a child psychiatrist), and then matching
as nearly as possible on the following variables, in order: another measure of severity of disturbance (a rating by an independent psychiatrist on a 4-point scale based on the mother's report), mother's mental state (rated by child psychiatrist on a 4-point scale following a semi-structured interview), and child's language development (similarly rated on a 3-point scale). This resulted in a group of children whose average rating of disturbance by the independent psychiatrist was significantly less severe than the treatment group. A second control group was then selected on the basis of matching to this variable alone. It is not reported whether the group thus formed differed significantly from the treatment group on the other variables used for matching on the first control group; only the data pertaining to the rating of disturbance by the independent psychiatrist is presented for the second control group. Additional clinician ratings were made of other family variables, such as father's mental state, the marital relationship, and "irritability to child by each parent" (p.351). Further data on severity of child disturbance, obtained from parents on the self-administered Behaviour Check-list (Richman, 1977) which contained the same items as the Behaviour Screening Questionnaire, was also available.

The progress achieved by each group was determined by statistical tests of the difference in mean scores between initial ratings and ratings after a year. The treatment group was reported to have attended an average of 41 half-day sessions over an average of 26 weeks. The main findings were that (a) significant improvement took place for all groups in the ratings of child disturbance, (b) significant improvement was not seen in any of the groups for ratings of family variables, and (c) no significant differences were found between the treatment and control groups with regard to average change in any of the ratings. The authors conceded that "there were few differences in outcome, and those that were found did not particularly favour either group" (p.349).

18 This suggests that 40 to 50 parents were separately seen by the single staff social worker mentioned in
However, the author’s conclusion can be questioned on the basis of a possible unaddressed confounding influence. A control group formed through selecting those with extreme scores from among the general population can be expected to show regression to the mean on re-test (Kidder, 1981). That is, the improvement in ratings of child disturbance in the control groups may be largely artifactual. A similar argument may apply to the ratings of mothers' mental state, and so forth, in the control group. While it would be difficult to argue that the treatment had a positive effect on the family variables since (a) no statistically significant improvement was seen in the treatment group, and (b) the amount of change evident in pre- and post-test mean scores appears to be negligible (standard deviations are not reported), the results with regard to child disturbance are quite ambiguous. Without a more adequate control group, it is not possible to assess the degree to which the reported improvement in child disturbance in the treatment group might in fact be attributable to treatment. With regard to addressing the clinical significance of the results, there are no details provided of the 4-point scale used for the psychiatrist's rating of severity of disturbance, and there are no discussions of post-test scores on the Behaviour Screening Questionnaire (Richman & Graham, 1971) or the Behaviour Check-list (Richman, 1977), two not very widely known tests, relative to community norms.

The same problem exists for the interpretation of the results of the five year follow-up on these children (Richman, Graham & Stevenson, 1983), where, again, the differences between the treated and untreated groups fail to reach statistical significance.

Anderson, Long, Leathers, Denny & Hilliard (1981) reported on the effectiveness of the therapeutic preschool of the Division of Child Psychiatry, Duke University Medical
Center, North Carolina, USA. It provided half-day sessions for children with mainly internalizing and externalizing problems, who attended "a regular day care center, preschool, Headstart program, or public kindergarten" (p.234) for the remainder of the day. Six 3 to 4-year-olds attended in the mornings and six 5 to 6-year-olds attended in the afternoons. The programme consisted of a therapeutic environment with a "constant, ongoing process in the classroom whereby teachers assist children whose behaviour is disruptive or withdrawn, and help them cope with encountered situations, crises, conflicts and feelings" as well as "ongoing parent work" (p.235).

The study involved 23 children (17 males & 6 females) whose average length of stay was 17 months. Pre-test and post-test ratings by teachers on the Preschool Behavior Questionnaire (Behar & Stringfield, 1974), a 30-item behaviour rating scale, provided comparison on four scores: Hostile/Aggressive Behavior, Anxious Behavior, Hyperactive/Distractible Behavior and Total Disturbed Behavior. In addition, three or four items particularly pertinent to each child were selected from the Questionnaire and rated weekly on a 4-point scale ranging from 1 (no concern) to 4 (great concern). The criterion for improvement here was "evidence for positive change in 50 percent of the problem areas" (p.236)\(^\)\(^{19}\).

The results were that (a) statistical tests showed a significant decrease between pre-test and post-test on all four scores of the Preschool Behavior Questionnaire, and (b) 17 of the total of 23 children (73%) fulfilled the criteria for positive change on their "target symptoms". The authors' main conclusion was that "overall, participation in the therapeutic preschool program resulted in positive changes in behavior" (p.239).

\(^{19}\) Additional analyses examining the relationship between outcome and the children's problem type which will be discussed in the next section [3.4].
The main problem with this conclusion pertains to the usual threats to internal validity in a design of this type. That is, in the absence of a no-treatment control group, it is not possible to rule out the possibility that the change in the treatment group may be an artifact of re-testing, or improvement which might have occurred without treatment, and so forth. The threat to internal validity presented by maturational effects is of particular concern in longitudinal studies of children; the present study involved 3 to 6-year olds seen over an average of 17 months. Another limitation is that the conclusion is based entirely on ratings performed by the teachers running the programme, who might be expected to have a special interest in demonstrating its effectiveness. They might also be expected to be vulnerable to the expectations and wishes of the researchers who were associated with the university department which, it is reported, provided them with day to day clinical supervision. Furthermore, there is a similar concern about how the criterion for improvement in "target symptoms" was applied; it is not reported how "positive change" was judged.

Another limitation of this study is that the outcomes were obtained at discharge only and not at follow-up. Not only does this result in no information with regard to the durability of the apparent improvement, post-tests conducted at the point of discharge may be capitalizing on random variation in the child's symptomatology. That is, it might be expected that one of the important factors determining the timing of discharge in programmes of this type would be symptomatic improvement in the child as observed by staff. A sceptical, Eysenck-style (1952) alternative interpretation of these results can be readily offered: the pre- to post-test improvements may be due to the children not being discharged from the programme until they show "spontaneous recovery". This would be an extreme view, and open to challenge on the basis of an examination of the preschool's policies, the actual circumstances under which
individual children came to leave the programme, and so forth. However, it does highlight the potential for inflation of the apparent effect of treatment when post-tests only occur at discharge from a treatment centre – a potential problem mostly ignored in the literature on evaluations of clinical treatments.

These problems warrant serious reservations about the authors' conclusion that participation in the program resulted in positive changes in children's behaviour. Furthermore, the amount of pre-treatment to post-treatment change, and the level of the children's post-treatment scores in relation to community norms, were not addressed. This makes it difficult for readers to make their own judgments about the clinical significance of the treatment effects, especially in view of the not very well known tests used.

Cohen, Bradley & Kolers (1987) reported on the effectiveness of the therapeutic preschool at Thistletown Regional Centre, Toronto, Canada, which provided children referred for internalizing and externalizing problems, global developmental delay and language disorders, with treatment consisting of half-day attendance, 5 days a week for one to two years. The description of the treatment modality is very similar to that of the other pre-school programmes which have been discussed: "Families were expected to observe and work with their child in the playroom under the guidance of staff and to maintain regular contact with their social worker" (p.688).

The study involved a comparison between a group of 53 children (85% male) who received treatment and a group of 45 children (56% male) recruited from community day-care centres as normal controls. A broad range of outcome measures were administered to the treatment group
at admission, 8 to 9 months later, and then again at discharge for those who continued treatment beyond a year. The control group was tested twice, 8 to 9 months apart.

The outcome measures were very comprehensive and consisted of the following.

(1) The Developmental Profile (Alpern & Boll, 1972), which involved gathering information from parents as well as direct testing on children, provided scores on five areas of child developmental ability: Physical, Self-Help, Social, Academic and Communication. These scores were averaged and then expressed as a developmental quotient (i.e. skill age/chronological age).

(2) The Reynell Developmental Language Scales (Reynell, 1969) which provides developmental quotients for expressive language ability and receptive language ability.

(3) The Draw-a-Line Slowly Test (Maccoby, Dowley, Hagen & Dagerman, 1965) was administered as a measure of motor impulsivity.

(4) The Smartie Delay Test, adapted from Golden, Montare & Bridger (1977), was administered as a measure of ability to delay gratification.

(5) The Child Behavior Checklist (Achenbach & Edelbrock, 1983) was completed by parents. The analysis involved three scores derived from the social competence items (Activities Score, Social Score and Total Competence Score) and three scores derived from the problem behaviour items (Internalizing Score, Externalizing Score and Total Problem Score).
(6) The Kohn-Rosman Problem Checklist and the Kohn-Rosman Social Competence Scale (Kohn & Rosman, 1973) are similar instruments but designed to be completed by teachers. It provides two problem scores, Apathy-Withdrawal and Anger-Defiance, and two competence scores, Interest-Participation and Cooperation-Compliance.

Additional ratings took place for the treatment group. At each of a series of regular three-monthly case conferences, a rating was made on the following instruments, based on the information available to the participants: (a) the Global Assessment Scale for Children (Sorrells, unpublished), where a child's general clinical functioning is given a score between 0 and 100, and (b) the Goal Attainment Scale (Kiresuk & Sherman, 1968), which rates progress on individualised clinical goals set for each child and family.

Separate analyses of data were performed for sub-groups of children who were globally developmentally delayed (N=12, 100% male), showed specific language delays (N=24, 92% male) or referred primarily for internalizing and externalizing problems (N=17, 64% male). The results of the third sub-group, which is most relevant to this review, were as follows.

There were no statistically significant changes seen in scores on general development or language ability, or on teacher ratings of problem behaviours or competence. Statistically significant improvement was reported on scores on the Smartie Delay Test between admission and discharge, but not between admission and the 8-9 months re-test²⁰. As for parental ratings statistically significant change was reported between admission and the 8-9 months re-test on the CBCL Activity score, and between admission and discharge on the CBCL Total Competence score. Furthermore, significant improvement was reported on all three CBCL

²⁰The results of statistical tests for the Draw-a-Line Slowly Test does not appears to be reported.
problem behaviour scores between admission and the 8-9 months re-test, and also between admission and discharge. However, "complete data were available for only 9 of 17 mothers" and hence "these results are likely to be biased by a subset of mothers who felt most positive toward the children and the program" (p.692).

The authors conclude that "gains were not observed among children who presented primarily with behaviour problems" (p.687) and that "our findings for behaviourally disordered children do not look promising" (p.692). Given that the nine children of the behavioural problem subgroup for whom data was available are reported to have shown a decrease in the Total Problem Score of the Child Behavior Checklist from an average T score of 67.60 (SD = 9.2) at enrolment to 60.45 (SD = 10.3) at discharge it is difficult to agree unreservedly with this conclusion. However, without a more sophisticated data analysis (e.g. comparison of the nine against the eight which dropped out), or more appropriate research design (e.g. a more relevant control group), it is difficult to draw any firm conclusion from this study.

3.22 Programs for infant school aged children

There is only one report on the effectiveness of a day treatment programme catering exclusively to infant school aged children.

La Vietes, Cohen, Reens & Ronall (1965) reported on the effectiveness of the Children's Day Treatment Center and School, New York City, USA. Contained within a mainly descriptive account of the programme and its achievements are some data on the proportion of children able to be re-integrated to a regular school. The programme appears to be very impressive in the intensity and duration of treatment (three year admission, one year outpatient follow-up and
further consultation available for a period of five years). However, re-integration to a regular school is quite limited as an indicator of treatment outcome.

Whether a child is maintained in a regular school reflects a complex social situation involving factors outside of the child, such as the expectations of the school with regard to conduct and academic achievement, the school's resources and ability to manage or tolerate the child's problem behaviours, the availability of alternative placement options remaining for the school, the follow-up support offered to the school, and so forth. Furthermore, there is no information provided as to how long after discharge this data was collected. It is difficult to meaningfully evaluate the reported results.

### 3.23 Programs for infant and primary school aged children

There are six reports on the effectiveness of day treatment programmes for infant and primary school aged children (Swan & Wood, 1976; Zimet et al., 1980; Grizenko & Sayegh, 1990; Grizenko, Papineu & Sayegh, 1993; Grizenko, 1997; Kotsopoulos, Walker, Beggs, Jones, 1996).

Swan & Wood (1976) reported on the effectiveness of the "developmental therapy" offered at the Rutland Center, Athens, Georgia, USA. An educationally orientated programme based on a specific theory of the stages of cognitive, emotional and social development, it catered to a very mixed population.

The 75 children (63% boys) in the study included a majority who might be broadly characterised as showing predominantly internalizing or externalizing problems (51 students), with others reported as carrying a primary diagnosis of organicity (7), "severe socioeconomic deprivation"
(7), psychosis (4), mental retardation (3), learning disability (2), or hearing loss (1). The average age of the children at entry was 8.7 years with a range from 4.9 years to 12.7 years. Average length of treatment was 23.9 weeks with a range from 10 to 60 weeks.

The reported outcomes were based on assessment forms developed within the developmental therapy programme and pertain to programme-specific goals. There are no references to behaviours or outcomes perceived by individuals outside the programme, or behaviours which take place outside the programme, or which take place after the child has completed the programme. Furthermore, there is no control group. With such limited internal and external validity it is difficult to come to any judgement about the reported results of this study. It is of interest, however, that satisfactory progression was generally reported except for a small group of 5 children who entered the programme at the highest level ("Stage Four"). This group - "in each instance they have been bright children about 12 years of age" (p.59) - showed deterioration. The amount and type of deterioration would appear to be greater than what can be attributed to regression to the mean based on random measurement errors. This issue of deterioration in academic performance amongst the most able in a programme will be further discussed below in relation to other studies with what appear to be similar findings.

Zimet, Farley, Silver, Herbert, Robb, Ekanger & Smith (1980) reported on the effectiveness of the psychoeducational day treatment center at the University of Colorado Health Sciences Center, Denver, USA. The programme ran five days a week for 11 months a year, with short vacations interspersed, and catered for up to 24 children and their parents at any one time. Aside from daily classroom activities, children received individual psychotherapy twice a week on average. Parental involvement in the programme "may include weekly individual, couple, and/or the total family group psychotherapy" (p.241). Average length of treatment was 2 years and 2
months with additional outpatient treatment if necessary. The types of problems shown by the children appear to have been internalizing or externalizing problems in the majority of cases, with a minority of 26% described as showing "developmental deviations" and a further 2% with psychotic disturbance.

The study involved the administration of a variety of instruments to a group of children\(^\text{21}\) (75% boys) at some or all of the following points in time: at entry into the programme, at termination, 3-6 months after termination, and 15-18 months after termination. Results from the following instruments were reported: (1) The Louisville Behavior Checklist (Miller, 1977a), consisting of 163 items pertaining to problem behaviours, was independently completed by each parent; (2) The School Behaviour Checklist (Miller, 1977b), consisting of 96 items pertaining to problem behaviours in the school setting, was completed by the child's teacher at the respective point in time; (3) The Wide Range Achievement Test (Jastak & Jastak, 1965), which tests ability in reading, spelling and arithmetic, was administered to each child; (4) The Wechsler Intelligence Scale for Children - Revised (Wechsler, 1974), a standard test of intellectual abilities, was administered to each child; and (5) The Piers-Harris Children's Self-Concept Scale (Piers & Harris, 1969), consisting of 80 true or false items pertaining to children's beliefs about themselves, was administered to each child.

Data analysis consisted of multiple \(t\) tests between the average scores on the various instruments at each point in time, with the following results.

1) On the Louisville Behavior Checklist, ratings by mothers and fathers are reported to show statistically significant decrease between the point of entry and (a) termination, (b) 3-6 months

\(^{21}\) The number of children in the study appears not to be reported.
follow-up, and (c) 15-18 months follow-up, on "most" of the 18 scales. It is also reported that significant changes were seen on "very few" scales between the point of termination and either follow-up point.

2) On the School Behaviour Checklist, a statistically significant decrease in the Total Disability score was reported to have taken place between the point of entry and (a) termination, (b) first follow-up, and (c) second follow-up, and also between termination and both follow-up points. The Academic Disability score, however, showed no statistically significant change over time.

3) On the Wide Range Achievement Test, a statistically significant increase in all three academic areas was reported to have taken place between the point of entry and (a) termination, and (b) second follow-up. (The WRAT was not administered at first follow-up). However, the average increase in ability between entry and termination expressed as grade levels is reported to have been at or below that which might be expected over a two year period.

4) The Wechsler Intelligence Scale for Children - Revised was administered at entry and at termination only, and a statistically significant increase is reported for Full Scale and Performance IQ scores, but not for Verbal IQ scores.

5) The Piers-Harris Children's Self-Concept Scale was also administered at entry and at termination only, and a statistically significant increase (i.e. indicating an increase in self-esteem) is reported on all subscales.

The authors conclude that they have demonstrated that "changes in school behavior, in academic performance, in IQ, in home behavior, and in self-concept did occur during the children's
enrollment at the Centre and at two follow-up points after termination", but that there is "no way to determine the extent to which changes would have occurred over this time period had the children remained in their previous setting or been placed elsewhere" (p.254). While this conclusion is largely consistent with the findings and adequately qualified in view of the study design (i.e. it is difficult to argue with the authors’ conclusion that no meaningful results about the effectiveness of the program can be drawn from statistical tests of significance on differences between pre-treatment and post-treatment scores in the absence of a control group), the claim of positive change in academic performance is not justified. This claim appears to be based on an argument that the children under treatment did not fall any further behind while "children left untreated would continue to fall further and further behind academically" (p.245) – such a claim is difficult to maintain in the absence of evidence from a control group or specific evidence cited from other studies. Furthermore, the authors misinterpret the statistical tests of significance - they fall into the common misunderstanding of the results of statistical tests of significance as denoting the strength of a finding. For example, in discussing the results on the Piers-Harris Children's Self-Concept Scale, p levels are used to rank subscale scores with regard to the relative size of the change achieved in treatment (p.253).

However, there is some attempt to address effect size and the level of post-treatment functioning relative to community norms. There are graphs illustrating mean standard scores at each data collection point, containing "pathological/normal" cut-offs. These indicate, for example (reading off Figures 4 & 5, pp. 252-253), a change in mothers' total score on the Louisville Behavior Checklist from a mean T-score of 79 at intake to 63 at termination, 64 at first follow-up, and 59 at second follow-up (approximately). This suggests a change from scores at intake indicating severe disturbance, to scores at termination and follow-up indicating levels of problem behaviours closer to the population norm and below the “pathological/normal” cut off (T score
of 65). Comparable results appear on fathers' ratings on the Louisville Behavior Checklist, and the Total Disability score of the School Behavior Checklist, but not the Academic Disability score which remained remarkably flat.

Hence, the authors' conclusion that positive change took place in the group of treated children over the course of treatment and was sustained at follow-up is not without support, except in relation to academic ability. In view of (a) the mean effect size, (b) mean post-treatment scores relative to community norms, and (c) the fact that gains were sustained at follow-up, the demonstrated impact of treatment on severity of problem behaviour appears to be quite substantial.

Grizenko & Sayegh (1990) reported on the effectiveness of the Lyall Preadolescent Day Treatment Program of the Douglas Hospital, Verdun, Quebec, Canada. The programme involved up to 20 children at any one time, with internalizing or externalizing problems, for an average attendance of seven months. Treatment was described as psychodynamically oriented, with the children receiving special education and attending therapeutic groups each day and the families attending "mainly systemic" family therapy each week.

The study involved the comparison of average scores at admission and at discharge for 23 children (91% male) on a wide variety of instruments. The instruments comprised the following: the Child Behavior Checklist and its parallel forms completed by parents, primary therapists (child care worker or nurse) and teachers; the Wide Range Achievement Test; several self-report

---

22 The graphic presentation of WISC-R results included mean sub-test scores but not, unfortunately, Full Scale, Verbal or Performance IQ scores. Nor are they reported in the text. The graph of results on the Piers-Harris Self-Concept indicates an increase from a mean standard score of around 50 to a mean standard score of around 62. The authors do not comment on what appears to be the unexpectedly high scores on this instrument at admission.
forms completed by the children - the Hare Self-Esteem Scale, the Index of Peer Relations, the Depression Self-Rating Scale, and the Hopelessness Scale for Children (Corcoran & Fischer, 1987); both forms of the Matson Evaluation of Social Skills with Youngsters (Matson, Rotatori & Helsel, 1983), available as a self-rating form for children and a form for completion by teachers; and the Family Assessment Measure (Skinner, Steinhauer & Santa-Barbara, 1983), a scale containing items pertaining to "task accomplishment, role performance, communication, affective expression, involvement, control, values and norms" (p.521) in the family, completed by the child and each parent separately.

Data analysis involved multiple paired t tests to compare pre- and post-test scores. It is reported that "discharge scores revealed a significant improvement on all standardised scales for behaviour (p<.0001), academics (p<.05), personality (p<.0001) and family measures (p<.05)" (p.521), leading the authors to conclude that "all children in the day treatment program improved significantly in behaviour, academic and personality functioning" (p.523). It is acknowledged that in the absence of a control group "the positive changes can not necessarily be attributed to the day treatment program" (p.524).

The data analytic approach is very similar to that of Zimet, Farley, Silver, Herbert, Robb, Ekanger & Smith (1980), discussed above, and while there are some small problems with the interpretation of results, there are similar indications that positive change had taken place. As to objections pertaining to the data analysis itself and the conclusion which the authors draw from it, there are, first of all, some relatively minor issues. The reported p value of .096 for the difference between average pre- and post-treatment ratings on the Family Assessment Measure by fathers is not consistent with the stated conclusion that significant improvement took place on all family measures at p < .05 (p. 521). Furthermore, it is difficult to understand the authors’
conclusion that "results of the study indicate that all children in the day treatment program improved significantly in behaviour, academic and personality functioning" (p. 523, emphasis added) other than as stating that each child showed a significant amount of improvement. This does not follow from the finding of statistically significant differences in the positive direction between average pre- and post-treatment scores. In addition to the common confusion of the statistical significance of results as denoting effect size, there is an inappropriate attribution of the finding about average scores to a conclusion about all individual children. While the conclusion may well be accurate, the data analysis reported does not address the issue of effect size for the treatment group, let alone show that a “significant” effect was seen in each individual case (which seems unlikely).

With regard to the evidence for meaningful change, however, an examination of Table 1 indicates a decrease in the mean standard Total Behavior Problem score on the Child Behavior Checklist from a T-score of 78.7 (SD = 6.6) to a T-score of 54.1 (SD = 11.2), which is a very impressive result. It is a large change, and the post-treatment score is very close to the population mean. A limitation of this study is that, unlike the study by Zimet, Farley, Silver, Herbert, Robb, Ekanger & Smith (1980), above, the post-test took place at discharge only.

Grizenko, Papineau & Sayegh (1993) then followed with a report on another 15 children attending the same day treatment centre. The research design differed in that a waiting-list control group of 15 children was included, and outcome data was gathered for the treatment group at 6 months post-discharge. The instruments used were as in the previous study except that ratings on the Child Behavior Checklist were obtained from parents only; ratings on the

---

23 While it is reported that the total sample of 30 children was 77% male, the gender breakdown is not reported for the two groups. Nor is a matching procedure based on gender reported.
Matson Evaluation of Social Skills with Youngsters were obtained from children only; and ratings on the Family Assessment Measure were obtained from one and not both parents\textsuperscript{24}.

The comparison between treatment and control groups involved the application of multivariate analyses of covariance to five groupings of dependent variables, with pre-treatment score and age at admission\textsuperscript{25} as covariates. Statistically significant differences between treatment and control groups are reported on parental ratings of behaviour problems (standardised Internalizing and Externalizing scores on the Child Behavior Checklist) and children's negative self-perceptions (scores on the Hare Self-Esteem Scale, the Depression Self-Rating Scale, and the Hopelessness Scale for Children). On the other hand, it is reported that no significant differences between treatment and control groups were found on children's self-report of the quality of their peer relationships (scores on the Index of Peer Relations and the Matson Evaluation of Social Skills with Youngsters), self-ratings of family functioning (child and parent ratings on the Family Assessment Measure), and academic performance (reading, spelling and arithmetic scores on the Wide Range Achievement Test).

The testing of change in the treatment group over time (intake, discharge and follow-up) involved the application of doubly multivariate analyses of variance to the five groupings of dependent variables. Parental ratings of behaviour problems, children's negative self-perceptions, children's self-report of the quality of their peer relationships, and self-ratings of family functioning all showed (a) significant multivariate $F$ with average scores changing in the direction of improvement, and (b) a significant difference between average scores at intake and

\textsuperscript{24} It is not specified whether this was the mother or the father. This may be relevant in view of the finding in their previous study suggesting that fathers may report less change than mothers on this instrument.

\textsuperscript{25} Age was identified as related to outcome in their previous study.
follow-up in the direction of improvement. However, there was no significant change in average scores over time on academic performance.

The authors conclude that "compared with a waiting list control group, multimodal treatment for a psychiatric population of children with disruptive behaviour disorders produces greater gains" and that "treatment gains were maintained at a 6-month follow-up" (p.133). This conclusion is supported by the demonstrated difference between treatment and control groups with regard to change in parental ratings of behaviour problems and children's negative self-perceptions. Furthermore, an examination of the unadjusted group means reported for children's self-report of the quality of their peer relationships and self-ratings of family functioning suggests that the failure to demonstrate statistically significant differences might be related to inadequate power associated with the small $N$ of this study. The results for these four groups of variables are consistent with the authors' conclusion that treatment had a beneficial effect.

The results for academic functioning, the fifth grouping of outcome variable, were different, however, and warrant further comment. In contrast to the other four groupings of scores, the mean pre- and post-treatment scores on the Wide Range Achievement Test are not suggestive of a difference between treatment and control groups. In fact, the pre- to post-test improvement in average scores on the WRAT seen in the earlier treatment group (Grizenko & Sayegh, 1990) was not repeated in the second study Grizenko, Papineau & Sayegh (1993) (see Table 1).
<table>
<thead>
<tr>
<th></th>
<th><strong>Pre-test</strong></th>
<th></th>
<th><strong>Post-test</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td><strong>Grizenko &amp; Sayegh (1990)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reading</td>
<td>72.3</td>
<td>(25.7)</td>
<td>80.4</td>
<td>(26.1)</td>
</tr>
<tr>
<td>spelling</td>
<td>72.0</td>
<td>(18.3)</td>
<td>78.1</td>
<td>(16.4)</td>
</tr>
<tr>
<td>arithmetic</td>
<td>71.2</td>
<td>(21.0)</td>
<td>84.5</td>
<td>(19.8)</td>
</tr>
<tr>
<td><strong>Grizenko, Papineau &amp; Sayegh (1993)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reading</td>
<td>81</td>
<td>(16.4)</td>
<td>79</td>
<td>(15)</td>
</tr>
<tr>
<td>spelling</td>
<td>75</td>
<td>(15)</td>
<td>75</td>
<td>(12.2)</td>
</tr>
<tr>
<td>arithmetic</td>
<td>85</td>
<td>(10.3)</td>
<td>83</td>
<td>(14.1)</td>
</tr>
</tbody>
</table>

**Table 1.** Pre- and post-treatment WRAT scores reported in Grizenko & Sayegh (1990) and Grizenko, Papineau & Sayegh (1993)

This difference is not discussed by the authors. Nor are there differences identifiable from their reports between the treatment groups in the two studies, or in the Lyall Preadolescent Day Treatment Program itself at the two points in time, which might account for the difference in outcomes. However, a hypothesis can be offered on the basis of the difference in pre-treatment scores between the two samples of children. The average pre-treatment scores for the earlier group lie much closer to the reported "problem range" of <70 than the later, while the post-treatment scores are very similar for both groups; it may be the case that the educational component of a treatment programme of this type, which emphasises emotional intervention over direct academic instruction and might also be expected to be conducted within a rather chaotic milieu, results in academic gains for the less able students but allows the more able students to fall behind\(^{26}\). This hypothesis can in principal be tested through a reanalysis of the

\(^{26}\) The reported standard deviations indicate a considerable range in the ability of the children.
data in subgroups. Moreover, there is some support for this explanation from the deterioration reported in the most able among the children in the evaluation of the educational therapy offered at the Rutland Center (Swan & Wood, 1976), as noted above, and the results of Kotsopoulos, Walker, Beggs & Jones B (1996) which is discussed below.

Grizenko (1997) followed up on 33 of the children from the two earlier studies, 5 years after discharge when their average age was 13 years (SD = 2.0). A further eight are reported as being lost to followup. It is argued that as “analyses revealed no differences between the subjects followed and those lost” with regard to clinical and family variables (including variables which in a later analysis of predictors of outcome was shown to account for 92% of the variance in outcome of the followup group) there is no selection bias. This neglects the fact, however, that of the 8 children lost to followup, two were “incarcerated at a youth detention facility” (p. 992) - it appears therefore that 2 children with possibly the worst outcomes were lost to follow-up.

It is reported that “repeated-measures analyses of variance showed that improvement was maintained on all measures between intake and 5-year follow-up” and that “children who were admitted to a day treatment setting appear to function well globally, even 5 years after discharge” (p. 989). However, the outcomes at 5 year follow-up on the Total Problem Score of the CBCL, with an average of 67.6 (T score, SD = 8.7), shows a significant amount of deterioration since discharge (T score = 56.3, SD = 7.5). The total problem score is, in fact, now in the clinical range (Achenbach, 1991). While this still remains well below the Total Problem Score for this subgroup at admission (T score = 81.7, SD = 6.4), the results suggest a more qualified success than indicated in Grizenko’s conclusions.
Kotsopoulos, Walker, Beggs & Jones (1996) reported on behavioural and academic outcomes for the day treatment program at the Royal Ottawa Hospital, Ontario, Canada. The program is described as a tertiary care facility that serves the greater Ottawa region, “accepting children with the most severe psychiatric disorders who cannot be managed by other services in the region” (p. 373). The types of problems shown by the children appear mainly to be externalizing, with a small minority of children showing internalizing problems. A maximum of 39 children receive academic instruction in classrooms conducted along behavioural lines, where “adequate social behaviour” is rewarded and “aggressive and defiant behaviour” leads to the withdrawal of privileges (p. 374). Cognitive-behavioural therapy for “impairment in social behaviour” is also conducted within the class setting, “distractibility and disruptive and aggressive behaviours are controlled with stimulant medication, often in combination with clonidine”, and children with language and visuo-motor problems receive individual or small group treatment from a speech therapist and/or an occupational therapist. It is reported that while the children are “seen by a child psychiatrist at regular intervals”, the “family counselling and parent management training” are conducted by a “child and youth worker” (p. 374). The average length of stay is 1.5 academic years.

The study involves a comparison of pre-treatment and post-treatment scores, on a range of behavioural and academic instruments, for 46 children (39 males, 7 females) admitted to the program. The behavioural instruments were the Child Behavior Checklist (Achenbach, 1991), the Teacher’s Report Form (Achenbach, 1991) and the Conners Teacher Questionnaire (Conners, 1969). The first instrument was completed by parents while the latter two were completed by program teachers. The academic instrument was the Kaufman Test of Educational Achievement (Kaufman & Kaufman, 1985), producing grade equivalent and age percentile scores for reading, spelling and maths.
Interestingly, the two teacher-rated problem behaviour instruments showed no statistically significant pre to post-treatment differences on any of the 8 subscores reported, while the parent-rated instrument (CBCL) showed a statistically significant change in the positive direction on both problem behaviour subscores reported (Internalizing Score and Externalizing Score). As for academic performance, the Kaufman Test showed a statistically significant difference in the positive direction on school grade equivalent scores in all three subjects, but no statistically significant differences on age percentile scores on any subject. The authors conclude that “this study showed that children who attended a day treatment and school program for approximately one academic year improved significantly in behaviour and academic performance” and that “the results of this study suggest that highly dysfunctional children presenting mainly with disruptive behaviour disorders may be successfully treated in day treatment programs” (p. 375).

This conclusion is not well supported by the results, however. The results of the teacher behaviour ratings, the parent behaviour ratings, and the academic testing need to be considered with care.

As for the teacher behaviour ratings, the authors suggest the lack of change in teacher ratings\(^\text{27}\) may be due to “low scores in the children’s initial assessment” (p.375). The authors are not suggesting a floor effect of some sort in the instrument as the pre-treatment Internalizing and Externalizing scores were in the “clinical range”\(^\text{28}\). They suggest, rather, that problem behaviours seen by Program teachers were lower than that exhibited in the child’s previous regular school, due to “low student to staff ratios, a highly structured program, and the early

\(^{27}\) For example, on the TRF, the mean Internalizing Score pre-treatment was 65.6 (T score, s.d. = 8.4) while the mean post-treatment score was 64.6 (T score, s.d. = 12.7); the mean Externalizing Score pre-treatment was 66.8 (T score, s.d. = 10.2) while the post-treatment score was 64.2 (T score, s.d. = 13.6).

\(^{28}\) Commencing at 63 (Achenbach, 1991).
initiation of pharmacotherapy” (p. 376) in the Program school. While it may well be the case that children’s behaviour in the Program class room is different from their behaviour in regular class rooms, this still does not explain the lack of change seen in their problem behaviour over the course of their treatment. It would still be expected that an effective treatment program, especially one incorporating a specific contingency management program to address class room behaviour, would reduce the level of problem behaviours from the “clinical range” to a level closer to the normal range.

As for the parent behaviour ratings, the change in the Internalizing Score of the CBCL from a pre-treatment mean of 68.9 (T score, s.d. = 10.6) to a post-treatment mean of 63.8 (T score, s.d. = 14.7), and the change in the Externalizing Score from a pre-treatment mean of 73.5 (T score, s.d. = 8.7) to a post-treatment mean of 66.3 (T score, s.d. = 14.7) are quite large. However, the post-treatment Externalizing score – representing the main problem area of the majority of these children and apparently the main focus of therapeutic efforts – remains in the “clinical range” of >63 (Achenbach, 1991). Given the weak internal validity of a pre-posttest design without a comparison group, and the fact that these results are prior to discharge from the Program (with the potential that holds for capitalising on “spontaneous recovery”, as described above [3.21]), the results provide limited evidence of therapeutic change.

As for the academic testing, the results on the Kaufman Test of Educational Achievement showed an improvement of approximately one year in grade equivalency, which is an unremarkable amount of change over approximately one academic year of treatment. As can be readily seen when expressed as percentiles (Table 2), the children’s performance, on average, showed very little change relative to their peers.
Notwithstanding, the authors do report an interesting finding with regard to academic outcomes: they report that gains in the reading score showed a significant negative association with pre-treatment scores on the Clinical Evaluation of Language Fundamentals –Revised (CELF-R) – a standard test of speech language pathology (Semel, Wiig & Secord, 1987). These results appear in Table 3, below.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>SD</td>
</tr>
<tr>
<td>reading</td>
<td>32.0</td>
<td>(29.3)</td>
</tr>
<tr>
<td>spelling</td>
<td>17.0</td>
<td>(20.6)</td>
</tr>
<tr>
<td>arithmetic</td>
<td>34.7</td>
<td>(25.0)</td>
</tr>
</tbody>
</table>

**Table 2.** Pre- and post-treatment scores on the Kaufman Test of Educational Achievement: Percentile of academic level (Kotsopoulos, Walker, Beggs & Jones, 1996)

<table>
<thead>
<tr>
<th></th>
<th>Receptive Language subscore</th>
<th>Expressive Language subscore</th>
<th>Total Language score</th>
</tr>
</thead>
<tbody>
<tr>
<td>change in Reading score</td>
<td><em>r</em> = -0.33</td>
<td><em>p</em> = 0.03</td>
<td>-0.19</td>
</tr>
</tbody>
</table>

**Table 3.** Correlations between change in pre and post treatment Reading scores (Kaufman Test of Educational Achievement) and pre-treatment CELF-R subscores and total score (Kotsopoulos, Walker, Beggs & Jones, 1996)

The negative association between change in Reading score and pre-treatment receptive language ability – indicating that children who are weaker in basic receptive language ability show more gain in Reading ability over the course of admission – is quite substantial (the *r* is “medium”
size by Cohen, 1965). The negative association between academic gain and pre-treatment academic ability appears to echo an interesting aspect of the results of Grizenko and colleagues (Grizenko & Sayegh, 1990; Grizenko, Papineau & Sayegh, 1993) and Swan & Wood (1976), noted above.

3.24 Programs for a broad range of age groups

There are two reports on the effectiveness of a day treatment programme which includes a broad range of age groups.

Halpern, Kissel & Gold (1978) report on long-term outcomes for a heterogeneous group of 85 children and adolescents discharged over a 10-year period (1964-1974) from the Day Treatment Unit of the Rochester Mental Health Center's Children and Youth Division, Rochester, New York. They were aged from 3 to 13 years, and included "psychotic, autistic, brain-damaged, learning-impaired, disinhibited, and assaultive types of children" (p.320). The treatment orientation appears to have been equally diverse: "the whole gamut of teaching modalities and socializing tactics that can be fit into the available timetable in a controlled fashion" (p.320).

The research method involved interviewing one of the child's parents on the phone and obtaining (a) a rating of the child's overall adjustment on a 5-point scale ranging from "very well-adjusted" to "very poorly adjusted", and (b) a rating of how much help the parent believed the child received from the Day Treatment Unit on a 5-point scale ranging from "helped a great deal" to "not at all". Data were also collected on whether the children were able to be maintained at a regular public school. If so, their teacher was asked to (a) complete a behaviour rating scale
(Gold & Reisman, 1971), (b) provide a rating of the child's overall adjustment in a format reportedly similar to that for parents, and (c) complete a "Personality Description Rating Scale".

The data analysis, the reporting of the results and the conclusions drawn in the report are quite confused. In the absence of a control group or pre-test scores, and considering the vagueness of most of the ratings asked of the participants and their vulnerability to investigator expectations and other biases, it is difficult to meaningfully evaluate the reported results.

Kiser, Millsap, Hickerson, Heston, Nunn, Pruitt, Rohr (1996) report on outcomes one year after discharge for a heterogeneous group of 114 children and adolescents aged between 5 and 18 (63% male) completing treatment over a 2½ year period at two unidentified programs associated with the Department of Psychiatry, University of Tennessee, Memphis, USA. The average length of stay was 3.8 months. Statistical tests are performed on data obtained at admission and at followup one year after discharge, including the Child Behaviour Checklist, with significant $p$ values interpreted very strongly.

“To demonstrate the value of our partial hospital service to others, there are splendid things to say based on the results. For instance, to the parent or consumer, we can say, ‘One year after treatment in our program, your child or adolescent will be having fewer problems overall, will most likely be making better grades in school, having fewer conduct problems, and will be a better friend’” (p. 88).

---

29 This instrument is not cited in the reference, and was presumably developed for this study; it is not mentioned again in the report.
The change in the Total Problem score of the CBCL, for example, has a $p$ value reported at .000. On the basis of the data analysis presented, the author’s conclusions are probably overstated given the weakness of a pre-posttest design without a comparison group. Yet, an examination of the reported Total Problem scores, 70.08 at admission and 63.14 at followup (T scores), suggests that non-trivial change might have been achieved, although the follow up score remains in the “clinical range” (Achenbach, 1991). That is, this is an example of a study where, on one hand, the conclusions are not justified by the data analysis provided, yet on the other hand, the data suggests that the treatment may have resulted in clinically significant changes for at least some of the children - changes which a more appropriate approach to the data may have revealed.
3.3 Studies pertaining to the effectiveness of residential programs

The studies in this section address with regard to residential programs the question: *how effective is the treatment under examination?*

As mentioned in the introduction to this chapter [3.1], there are fewer reports on the effectiveness of residential psychiatric treatment for children and adolescents than of day treatment. This section is organized in a similar way to the preceding section.

3.31 Programs for infant and primary school aged children

One study pertains to a treatment program for infant and primary school aged children.

Winsberg, Bialer, Kupietz, Botti & Balka (1980) reported on a comparison between the effectiveness of residential treatment provided by **Kings County Hospital**, Brooklyn, New York, USA, and a home-based community treatment programme. The study involved a group of 49 children with a typical history of "chronic, severe behaviour disorders" (p.414), ranging in age from 5 years 3 months to 13 years 2 months, randomly assigned to either treatment. The externalizing problems were associated with brain damage in a few of the cases.

The children in hospital (24 children, 79% male) received "individually designed treatments, including pharmacotherapy, psychotherapy, and milieu therapy" and attended the hospital school (p.416). Their families were offered "social casework, counselling and psychotherapy" (p.416). The duration of treatment was two months for 5 children, four months for 7 children, and six months for 12 children. The children undergoing community treatment (25 children,
88% male) received pharmacotherapy but not "traditional psychotherapeutic services"\textsuperscript{30} and attended community schools. It is reported that "treatment emphasised support for the parents and the active recruitment of social services for the family" and that "the distinguishing characteristics of the community-care program were continuous availability of staff, persistent advocacy, and treatment flexibility" (p.415). Duration of treatment was four months for 2 children, and six months for the remaining 23 children.

Data were gathered during an observational period in hospital prior to random allocation to either treatment group, and then again after treatment, on the following instruments.

1) The Conners Behavior Rating Scale (Conners, 1969), consisting of 39 items pertaining to problem behaviours, and The Devereaux Elementary School Behavior Rating Scale (Spivack & Swift, 1967), consisting of 47 items pertaining to problem behaviours in a class setting, were completed pre-treatment for all children by the Hospital teacher as well as their teacher in the community. Post-treatment ratings were obtained from Hospital teachers for the Hospital-treatment group and from community teachers for the community-treatment group.

2) The Devereaux Child Behavior Rating Scale (Spivack & Spotts, 1966), consisting of items pertaining to problem behaviours in general social settings, was completed pre-treatment for all children by a member of the Hospital staff as well as the parents. Post-treatment ratings were obtained from Hospital staff for the Hospital-treatment group and from parents for the community-treatment group.

\textsuperscript{30} It is reported, however, that in three cases, where it was thought to be required, referrals were made to community mental health clinics.
3) Sub-tests from the Metropolitan Achievement Test (Durost, Bixler, Wrightstone, et al, 1970) pertaining to reading and mathematical abilities were administered to the children. For non-readers, Level 1 of the Stanford Early School Achievement Test (Madden & Gardner, 1969) was substituted.

4) The Psychiatric Status Schedule (Spitzer, Endicott & Cohen, 1968), an interview-based evaluation of psychiatric symptomatology and "role functioning" in adults, was administered to the mothers.

5) The Family Functioning Checklist, developed for this study, was completed by a caseworker. It contains items pertaining to four aspects of family adjustment: child care, family interaction, financial conditions, and housekeeping.

The results were as follows. With regard to problem behaviours in class, the community group showed statistically significant pre- to post-treatment reduction in average ratings by community teachers on three sub-scores of the Conners Behavior Rating Scale and the Devereaux Elementary School Behavior Rating Scale. In contrast, the Hospital group showed a statistically significant reduction in average ratings by Hospital teachers on only the Inattentiveness sub-score of the Conners Behavior Rating Scale. With regard to problem behaviours outside the classroom, the community group showed a statistically significant reduction in average ratings by parents on the Devereaux Child Behavior Rating Scale, while the Hospital group showed no statistically significant change in Hospital staff ratings on the same instrument. The results were more similar for the two groups on the other outcome scores. With regard to academic ability,

31 Only the Aggressivity, Inattentiveness and Hyperactivity scores were examined, with the Anxiety and Sociability scores left aside.
both groups showed significant improvement in average scores for reading and arithmetic. Neither group showed significant change in average scores with regard to mothers' psychiatric symptomatology and "role functioning" on the Psychiatric Status Schedule, or on ratings of family functioning on the Family Functioning Checklist.

The authors admit to some difficulties in the interpretation of the results due to the lack of concurrence in ratings of problem behaviours between community and Hospital teachers, and between parents and Hospital staff. As discussed above [2.52], it is difficult to know to what degree these results reflect different implicit norms in the raters or actual differences in the child's behaviour in different settings. The authors come to few conclusions beyond stating that the data "indicate improvement in both problem behaviours and educational achievement in the community group in both home and school settings" (p.418), although the change in average academic abilities for either group is recognised to be a "trivial" (p.417) amount. (The analysis does not appear to take into account the improvement in abilities normally expected in a student over time). They make little comment on the results for the Hospital group, but with (a) the trivial result for academic abilities, and (b) the very small amount of change apparent in average scores on the Conners Behavior Rating Scale (including the Inattentiveness score), the impact of residential treatment provided by the Hospital cannot be said to have been, on average, very substantial.

3.32 Programs for high school aged adolescents

There are two reports on residential treatment programmes for high school aged adolescents (Warren, 1965; Blackman, Eustace & Chowdhury, 1991).
Warren (1965) reported on a long-term follow-up of 157 adolescents admitted for residential treatment at the Adolescent Unit, Bethlem Royal Hospital, London, UK. While various conclusions are presented with regard to outcomes for particular diagnostic groups, given (a) the absence of a control group, (b) the length of the follow-up (at least 6 years post-discharge), (c) the heterogeneity of the population, and (d) the emphasis on classification according to psychiatric diagnoses rather than data pertaining to change in dimensions of severity of disturbance (i.e. emphasis on type rather than severity of disturbance), this report represents an epidemiological study on the later course of adolescent psychiatric disturbance rather than an evaluation of treatment.

Blackman, Eustace & Chowdhury (1991) reported on a one to three year follow-up of adolescents admitted to CASE (Child and Adolescent Services) House, Edmonton, Alberta, Canada, a "ten-bed, free standing semi-secure residential treatment program for severely dysfunctional adolescents with serious psychiatric problems" (p.473). They describe the program as being “specifically designed to address the problems of the adolescent who had previously been exposed to many unsuccessful treatments, had multiple problems and little community or family support. These children have often had extensive involvement with child welfare authorities, several unsuccessful placements, and are considered to be high-risk cases" (p.473). Apart from severe internalizing and externalizing problems, the adolescents were reported to show psychotic, affective, or eating disturbance. The treatment philosophy is based on the conceptualization of the residential unit as a therapeutic family with "daily insight-directed community therapy group and other structured and unstructured small therapy groups employing a mainly psychodynamic, systems-oriented framework" (p.473). Family therapy, educational remediation, parent education and support, and medication appear to play ancillary roles.
The study involved "two trained raters who were not involved in the treatment program" who made ratings at three points in time (admission, discharge, and one to three years after discharge) based on data gathered by "an independent research team" (p.475) on 40 individuals (50% male) admitted to the programme. Of these, 6 were unable to be located after discharge and were therefore lost to follow-up. Ratings were obtained on the following instruments:

1) The Global Assessment Scale (Endicott, Spitzer, Fleiss et al., 1976), where an individual's general clinical functioning is given a score between 0 (dysfunctional) and 100 (healthy).

2) The Level of Functioning Scale (Newman & Rinkus, 1978), which involves a rating on a 1 (severe pathology) to 9 (functions well) scale of an individual's functioning across four areas: physical functioning, interpersonal relationships, social role performance, and psychological signs and symptoms.

3) The Adolescent Functioning Scale, an additional instrument developed for this study, which involves a rating on a 1 (no problems) to 5 (severe problems) scale of an individual's functioning in each of seven areas: socialization/peer relationships, family functioning, school functioning, self-destructive activities (suicide), personal (somatic) care, acting out, and substance abuse.

Average score on the Global Assessment Scale at admission was 28.2 which indicates "very severe pathology... these adolescents were unable to function in almost all areas of life or their behaviour at the time was influenced by either delusions or hallucinations" (p.476). At discharge, following an average admission of 4.9 months, the average GAS score was 48.8, and at follow-up it was 68.8, representing a significant amount of continued clinical improvement.
The changes in scores between admission and discharge, and between discharge and follow-up, were both reported to be statistically significant on the Wilcoxon Matched Pairs Signed-Rank Test. The results for the Level of Functioning Scale were similar, with a change in average score from 2.5 at admission to 4.4 at discharge and 7.1 at follow-up (statistically significant change in scores between admission and discharge and between admission and follow-up).

In addition to the average score, the presentation of data pertaining to the Adolescent Functioning Scale included the proportion of cases showing "serious" or "severe" ratings in each of the seven areas at admission, discharge and follow-up. They were, respectively: 58%, 8% and 3% for socialization/peer relationships; 76%, 35% and 6% for family functioning; 59%, 3% and 0% for school functioning; 17%, 0% and 0% for self-destructive activities (suicide); 21%, 4% and 0% for personal (somatic) care; 67%, 12% and 3% for acting out; and 29%, 3% and 0% for substance abuse.

The authors conclude that "a significant improvement was found in the adolescent's level of functioning" and that "these treatment gains were maintained at the time of long term follow-up" (p.472). While the tests of significance performed on change in average scores over time are not very meaningful in the absence of a no-treatment control group, the reporting of treatment effect size and level of post-treatment functioning leaves little doubt that clinically significant change was achieved. The major reservation is that the potential for bias, if any, due to the special interests of the investigators is hard to evaluate as the source of primary data on which the ratings were based (the "two trained raters who were not involved in the treatment program" who made their ratings based on data gathered by "an independent research team") is not entirely clear. There is also the potential systematic bias in the follow-up group due to the possibility that the adolescents unlocatable at follow-up may be those experiencing the worst
outcomes, although at only 15% it is unlikely to have a very large influence on these results. On the basis of the large amount of sustained clinical change demonstrated in a large proportion of this group of adolescents with severe and long-term difficulties previously unresponsive to treatment, the demonstrated effect of this treatment programme is very impressive.

3.33 Programs for a broad range of age groups

There are two reports on the effectiveness of residential treatment programmes for a broad range of age groups (Levy, 1969; Treffert, 1969), both located in the United States. There is enormous contrast between the two programmes in available resources and treatment philosophy; one is a prestigious private residential facility emphasising intensive long-term psychotherapy, and the other a ward in a state mental hospital which emphasises a "predictable, consistent milieu" (Treffert, 1969, p.746).

Levy (1969) reported on an attempt at long-term follow-up on 113 children and adolescents aged between 5 and 15 years who were admitted for residential treatment at the Children's Hospital of the Menninger Clinic, Topeka, Kansas, USA. The programme is described as "psychoanalytically oriented, with a high staff-patient ratio and a full range of care, school, and treatment resources, including intensive psychotherapy" (p.1633). Treatment was mostly long-term, with 15% of the group admitted for less than 6 months, 70% from 6 months to three years, and 15% from three years to ten years. The majority appeared to show severe internalizing and externalizing problems, with another 25% psychotic and 10% organic.

There is no mention of direct clinical involvement with the family, and it might be assumed that many of the children and adolescents were separated from their families by considerable
geographical distances. In fact, it is reported that at the end of treatment "about half went home (some with a recommendation for further treatment); a little more than one-fourth remained in town for foster care homes and continued in psychotherapy; about one-fourth were transferred to other settings for continued treatment or care; and two left on their own" (p.1634). Considered together with what is reported about the lengths of admission, this suggests that for a considerable proportion of the children and adolescents the Hospital became home; the replacement of parents by Hospital staff is most striking in the proportion of patients who did not return home but entered foster care to remain in proximity to the Hospital and continue outpatient therapy.\footnote{Some of the most interesting aspects of the treatment are implicit in the arrangement but not directly addressed in the report. An example of this is that while it might be expected that the families were very unusual with regard to socio-economic status, no socio-demographic data are reported.}

For the study, the parents and guardians or associated mental health professionals, and in some cases the former patients themselves, were contacted and interviewed by telephone or letter. There were 100 individuals for whom data were able to be collected; their age was between 17 and 41 years and they had been discharged from the Hospital 8 to 23 years previously. While the descriptions based on the author's subjective impressions and the comments he received from the respondents are of considerable interest, as with Warren's (1965) long-term follow-up of adolescents discharged from the Adolescent Unit of Bethlem Royal Hospital mentioned above, the lack of structure or design means that the data collected does not permit an evaluation of the effectiveness of Hospital admission to take place.

Treffert (1969) reported on a five-year follow-up of 176 individuals admitted to the \textbf{child-adolescent unit at Winnebago State Hospital}, Winnebago, Wisconsin, USA, between July 1962 and December 1967. The unit is described in the following way: "The Winnebago state
hospital program directs itself toward the child who is mentally ill and who requires a period of
treatment in a relatively closed setting in a program relying heavily on psychiatrically
supervised structure and milieu. This is different from residential treatment with heavy emphasis
on individual psychotherapy, and is different also from residential treatment with casework
orientation" (p.745). There is an emphasis instead on the "attempt to orient, indoctrinate and
institute" the principles of a "predictable, consistent milieu" (p.746). No clinical work involving
the family appears to have taken place.

Of the original group of 176 identified for follow-up, eighty seven\textsuperscript{33} were interviewed along
with a parent or guardian. It is unclear how 50\% came to be lost to follow-up but one criterion
for inclusion appears to have been discharge for at least 12 months. The former patient and their
parent completed a questionnaire, presumably designed specially for this study, which included
(a) selecting from a list of 13 items "the five most helpful factors in hospitalization" (p.751), and
(b) making retrospective ratings of the patient's improvement over the course of admission, and
since discharge, on a seven point scale: -3 (got much worse), -2 (considerably worse), -1
(slightly worse), 0 (stayed the same), +1 (improved slightly), +2 (improved considerably), and
+3 (improved a great deal).

The quality of the Hospital experience is hinted at in the result, reported without irony, that both
the patients and their parents rated "desire to leave hospital" as the number one most "helpful
factor of hospitalization". It is also reported that "some children and some parents felt there
should have been more direct contact between social workers, doctors, and the patient" and that
"parents felt the hospital could have been more diligent about communicating to them the child's
progress and status and in general wished for better communication between the hospital staff
and themselves" (p.752). With regard to the ratings about improvement, it was reported that "52 patients (80%) felt they had improved while hospitalized, nine (14%) felt they had remained the same, and four (6%) felt they had gotten worse" with an average rating of +1.23. Self-ratings of improvement since discharge were reported to be an average of +1.56. Parents' ratings of improvement during hospitalization were reported to show even more improvement, with 80% indicating improvement (average rating of +1.4), and 77% indicating improvement since discharge (+1.6).

The authors conclude that "a clear majority had improved, [and] had maintained their improvement after discharge" (p.752).

There is a serious problem pertaining to the quality of data obtainable with an unstructured global rating of this sort, especially with regard to its vulnerability to the expectations and special interests of the participants. It is difficult to draw any meaningful conclusions from the reported results.

---

33 Breakdown by gender is not reported for this group, although it is reported elsewhere that males constituted about 75% of admissions.
34 c.f. McCord's (1978) classic evaluation of the Cambridge-Somerville, as described.
3.4 Studies pertaining to predictors of outcome

The studies considered in this section address the question: *what factors or variables account for individual differences in outcome?* As outlined [2.21], the design requirements for this question are quite different from the design requirements for addressing the question of the effectiveness of treatment.

A handful among the studies examined in the previous two sections [3.2, 3.3] also addressed the question of what accounts for the differences in outcome within the populations who participated in the various treatments.

3.41 Child variables

Perhaps surprisingly, only one study reported on the influence of *gender* on outcomes. The report by Eisenberg, Lansdowne, Wilner & Imber (1962) on the outcomes for a sample of preschooler attending the Children's Guild, Baltimore, USA [3.21] concluded that girls showed a "greater degree of recovery" (p.23) than boys. Given the problems in the interpretation of the results, as outlined, this conclusion cannot be accepted.

The study by Anderson, Long, Leathers, Denny & Hilliard (1981) of the therapeutic preschool associated with the Division of Child Psychiatry, Duke University Medical Center, North Carolina, USA, [3.21] was concerned with the relationship between outcome and the child's *problem type*. Each child was classified as "hostile/aggressive" or "anxious" (ie. externalizing or internalizing) on the basis of their scores on the Preschool Behavior Questionnaire. Only 50% of the children among the "hostile/aggressive" group (n=10) reached the criteria for positive
change in target symptoms, compared with 92% among the "anxious" group (n=13). This is reported to give a \( p \) value close to .03 on the Fisher Exact Test.

The problems identified in this particular study with regard to (a) the design of the study, (b) the potential for a general bias in the teacher-raters towards seeing improvement in their students, and (c) the potential capitalization on random variation in the child's symptomatology due to the post-test taking place at discharge, relate to conclusions about the overall effectiveness of treatment. The same objections do not arise in relation to the authors' conclusion here that "there was a significantly greater probability that the anxious children would benefit more from the program than the hostile/aggressive children in terms of a reduction in problem behaviours" (p.238). That is, while there is reason to be concerned about the meaningfulness of the reported outcomes in the absence of relevant control groups, and reason to suspect the operation of factors likely to bias the results towards augmenting the amount of change, there is no reason to suspect that these factors operate differently for preschoolers showing internalizing as opposed to externalizing behaviours. Limitations remain with regard to the very small size of the study and the lack of follow-up (i.e. the durability of the treatment effects), but it is not unreasonable to conclude that that this therapeutic preschool, at least during the period of this study, was less effective with the externalizing type than the internalising type of child problems.

The studies which emphasise \textbf{psychiatric diagnosis} as a predictor of outcome - such as the longer term follow-up studies by Warren (1965) of adolescents 6 years after discharge from the Adolescent Unit, Bethlem Royal Hospital, London, UK \cite{3.32}, or by Levy (1969) of 5 to 15 year olds followed up 8 to 23 years after discharge from the Children's Hospital of the Menninger Clinic, Topeka, Kansas, USA \cite{3.33} – cannot be said to represent an evaluation of treatment, as discussed.
There were two studies which did not directly address the issue of academic abilities as predictors of outcome but, as already noted, contain relevant findings. Swan & Wood's (1976) study on the effectiveness of "developmental therapy" at the Rutland Center, Athens, Georgia, USA, [3.23] found that the group of children who entered the programme with the highest level of academic ability showed deterioration in this area over the course of treatment. Furthermore, an examination of data reported in Grizenko & Sayegh (1990) and Grizenko, Papineau & Sayegh (1993) [3.23] on the Wide Range Achievement Test suggests that among the emotionally and behaviourally disturbed infant and primary school children attending the Lyall Preadolescent Day Treatment Program of the Douglas Hospital, Verdun, Quebec, Canada the less able students studied improved in academic ability while the relatively more able students studied did not. Furthermore, there was a study which directly addressed the issue of academic abilities as a predictor of academic outcome. Kotsopoulos, Walker, Beggs & Jones (1996) showed in their study of a day treatment program [3.23] that children who are weaker in basic receptive language ability showed more gain in Reading ability.

There is an interesting finding by Grizenko (1997) with regard to the association between children experiencing problems during their mother’s pregnancy as assessed by “the intake assessment, medical and obstetric records” (p. 993) and outcomes 5 years after discharge from day treatment. This may reflect the correspondence between prenatal injury and problem behaviours in later childhood rather than an aspect of response to treatment. On the other hand, the “problem pregnancy” variable may reflect problems in the early attachment between mother and baby, or perhaps a complex interaction between organic and attachment factors. In any case this is a finding worthy of further study and an aspect of children’s early development neglected
in most studies of the impact of treatment on severe psychological and behavioural problems in childhood.

3.42 Family variables

Cohen, Kolers & Bradley (1987), in a companion article to Cohen, Bradley & Kolers (1987) discussed above [3.21], examined predictors of outcome among the heterogeneous members of the treatment group in their study of the therapeutic preschool at Thistletown Regional Centre, Toronto, Canada. Among these predictors was an index of psychosocial risk, apparently developed for this study, which consisted of items pertaining to family socio-economic status, family structure (i.e. double vs. single parent), marital disharmony and maternal depression. However, as noted, the group was particularly heterogenous, including children showing global developmental delay and/or specific language delay and/or internalizing and externalizing problem behaviours. The multiple regression analyses of the relation between predictors and outcome did not involve subgrouping into the three main diagnostic categories, nor were the diagnostic categories included among the predictors. It is therefore difficult to understand the results for the population of interest in this present review.

Goldfarb, Goldfarb & Pollack's (1966) study of 26 psychotic infant and primary school aged children in day or residential treatment at the Ittelson Center for Child Research, Bronx, New York will be discussed in detail in the next subsection on treatment variables [3.43]. It is briefly mentioned here that they report an interesting interaction between diagnostic sub-type and treatment type which they suggest indicates the disturbing influence of family on outcomes for children with "nonorganic" schizophrenic disturbance.
3.43 Treatment variables

Studies of treatment variables as predictors of outcome for severely disturbed children and adolescent undergoing intensive day or residential treatment are surprisingly rare. **Day versus residential treatment** will be discussed in relation to the study by Goldfarb, Goldfarb & Pollack (1966) [3.42], and **therapeutic alliance** will be discussed in relation to Grizenko (1997) [3.23]. The important issue of the effectiveness of **specific treatment types** – which is central to the “empirically-validated” treatment movement - will be considered in relation to the influential study by Kazdin, Esveldt-Dawson, French & Unis (1987) yet to be discussed.

Goldfarb, Goldfarb & Pollack (1966) reported on a comparison between day and residential treatment for infant and primary school aged children showing schizophrenic disturbance at the **Ittelson Center for Child Research**, Bronx, New York, USA. The study involved 26 children allocated from a single waiting list to either day or residential treatment on the basis of vacancies as they occurred, while matching with regard to age (within one year), organicity (positive versus negative "neurological findings" in the independent opinion of a child neurologist), and IQ (within the following classes formed on the basis of the Full IQ score on the WISC: non-scorable, 46<FIQ<69, 70<FIQ<89, and 90<FIQ). This resulted in three diagnostic subgroups: three matched pairs of children "so low in ego" (p.122) as to be non-scorable on the WISC (2 pairs of boys, one pair of girls), six matched pairs of "organic" children (all boys), and four matched pairs of "non-organic" children (two boys, two girls).

Outcomes were assessed on the Scale for the Psychiatric Appraisal of Ego Status, a 5-point rating scale (1 "very severely impaired", 2 "severely impaired", 3 "moderately impaired", 4 "mildly impaired", 5 "normal") developed for this study. The scale includes clinical descriptions
of each step pertaining to the ability to differentiate important persons, communicate with others, care for self, and so forth; the authors report that "the broad gaps between scale steps enhanced the certainty that a change of even one step interval represented significant clinical shift" (p.122). The rating was performed by a psychiatrist at admission and then annually for three years. Additional outcome data were obtained at the same time on the reading sub-test of the Metropolitan Achievement Test and on the WISC.

The results were as follows. Among the three "non-scorable" pairs, all remained at level 1 on the Scale for the Psychiatric Appraisal of Ego Status, all remained non-readers and all remained unscorable on the WISC. Among the six "organic" pairs, 4 of the day status children showed improvement and 5 of the residential status children showed improvement (defined as increase of at least one step on the Scale for the Psychiatric Appraisal of Ego Status). Average reading ability after three years was very similar for the day and residential subgroups, but residents had a lower average score on admission, and their average change score was significantly greater than the average change score for day children (statistically significant at the .05 level on a paired $t$ test). The authors dismiss this, however, as an artefact associated with the score allocated to non-readers at admission. Average improvement in IQ was very similar for day and residential subgroups.

However, among the four "non-organic" pairs, none of the day status children showed improvement and 3 of the residential status children showed improvement - "the hypothesis of no difference between day and residential children is rejected at the 0.07 level" (p.125). Average reading ability at admission was very similar for the day and residential subgroups, but residents had a higher average score after three years and their average change score was greater than the average change score for day children to a statistically significant degree. Similarly, while IQ
scores at admission were very similar, residents had a higher average score after three years (103 vs 92) and their average change score (16 points) was greater than the average change score for day children (4 points) to a statistically significant degree.

The authors conclude that "schizophrenic children who on admission were unscorable on the Wechsler Intelligence Scale for Children, and were the most severely impaired in ego status (so that they lacked language, could not discriminate important and other persons, showed no contacting behavior to humans, were extremely aberrant in receptor response, and were entirely incapable of self-care), showed no significant improvement in either day or residence. Among the children who were scorable on the WISC, the organic children in day treatment showed progress that was not different from that of the organic children in residential treatment. Finally, the nonorganic children in residence gave evidence of more improvement than did the matched children in day treatment" (p.127).

As the authors point out, while the sub-groupings made the already small numbers in the study even smaller, clinical homogeneity was achieved which permitted meaningful analysis of data. For example, the results pertaining to the difference between day and residential groups among the "nonorganics" are significant with regard to effect size as well as statistical criteria. A limitation to the study was that there was no follow-up after the children were discharged. This is particularly relevant in view of the hypothesis which the authors offer in explanation of the observed interaction between treatment type and schizophrenia type. They suggest that "we do know that psychosocial factors in the family are a more primarily disturbing influence in the nonorganic group than in the organic group. It may well be, therefore, that the more total separation of the nonorganic child from disordered family influence offered by residential
treatment accounts for the greater effectiveness of the residential over the day program for this child" (p. 127).

The results presented and the interpretation which is put forward are, as the authors suggest, "provocative". Despite the obvious limitations of the study arising from its very small numbers, and the assessment being conducted by a psychiatrist who was (presumably) not blind to the treatment conditions nor the theoretical assumptions of the treatment facility, the analysis of outcome data in clinically meaningful subgroups has provided clinically meaningful results for this special population (i.e. those diagnosed as suffering schizophrenia, which is a very rare condition in childhood) that warrant further study.

Grizenko (1997), discussed above [3.23], provides interesting results with regard to the association between therapeutic alliance and 5 year outcomes. It is reported that a stepwise multiple regression analysis was used to examine predictors of CBCL Total Problem Scores at 5 year follow-up. The predictor variables included “parental cooperation” based on their attendance at scheduled family therapy session\(^\text{35}\), CBCL Externalizing Score at admission, CBCL Total Problem Score at admission, problems during pregnancy (as discussed above [3.41]), age at admission, and “history of child neglect” as assessed by “the intake assessment, outpatient records, and social service records” (p. 993). It is reported that a very large amount of the variance in outcome \((r^2 = .92)\) was explained by four predictors - in order of entry, “parental cooperation”, CBCL Externalizing Score at admission, CBCL Total Problem Score at admission, and “problem pregnancy”. While the correlation between CBCL scores at admission

\(^{35}\) Assessed on a 3 point scale: 1 = parents came to more than 75% of all scheduled family therapy sessions, 2 = parents came to 50% to 75% of all schedules family therapy sessions, and 3 = parents came to fewer than 50% of all scheduled family therapy sessions.
and CBCL sores at followup is unremarkable, the finding with regards to “parental cooperation” and “problem pregnancy” is of interest.

The association between “parental cooperation” and outcomes is a very strong one with an adjusted $r^2$ of .55 – that is, a remarkable 55 percent of the adjusted variance was accounted for by this factor. Considering the way in which this variable was constructed (i.e. the proportion of parental attendance at scheduled family therapy sessions), “parental cooperation” might also be called “therapeutic alliance” - the clinically relevant concept of an active working partnership between therapist and client now empirically well supported as an important general factor underlying therapeutic success across many different treatment modalities (Henry, Strupp, Schact & Gaston, 1994; Orlinsky, Grawe & Parks, 1994). The degree to which parents are willing to submit themselves to the gruelling task of examining their own contribution to their child’s problems may also reflect the quality of the relationship between child and parents, the quality of the relationship between parents (in two parent families), and personal qualities such as emotional maturity in the parents. Most likely, it might be expected to reflect a complex interaction between these factors. This also is a finding worthy of further study.

Specific claims for empirically validated treatment types, when they do appear in reviews of the literature in this area, do not stand close scrutiny. For example, Blanz & Schmidt (2000) state that: “..it can cautiously be concluded that psychiatric hospitalization of children and adolescents is often beneficial, particularly if special aspects of treatment are fulfilled (e.g. good therapeutic alliance, treatment with a cognitive-based problem-solving skills training package, or planned discharge)” (emphasis added, p. 703). This summary stands in contrast to the contemporaneous review of the same literature by Fonagy, Target, Cottrell, Phillips & Kurtz (2002) – first discussed in Chapter 1, then again in 2.42 - which concludes that “specialist
residential or day care facilities” needs to be regarded as an area within child and adolescent mental health with “a startling and worrying lack of evidence to guide practice (p. 390)”.

However, in examining Blanz & Schmidt (2000) for the evidence leading to their confident and specific claims with regard to the effectiveness of a cognitive-based problem-solving skills training package, it is possible only to find a reference to the review by Pfeiffer & Strzelecki (1990). Turning to Pfeiffer & Strzelecki (1990), however, these authors report only that their review “identified four studies that looked at the impact of various aspects of the planned interventions: therapeutic alliance (Clarkin et al, 1987), planned discharge (White et al, 1979), completion of a treatment program (Gossett et al, 1977), and the efficacy of a cognitive-based problem-solving skills training package (Kazdin et al., 1987). In all instances, the investigators found that the specific treatment variables correlated with positive outcomes…” (p. 851). Blanz & Schmidt’s (2000) uncritical, unqualified, decontextualized extraction of this statement by Pfeiffer & Strzelecki (1990), and placement in the abstract of their own review, is misleading. It suggests more solid evidence for the four purported predictor variables than the one study per variable found in Pfeiffer & Strzelecki (1990).

Regardless, the Kazdin, Esveldt-Dawson, French & Unis (1987) study demonstrating the efficacy of cognitive-behavioural problem solving skills training is worthy of closer examination. The study involved 56 children (80% boys) selected from among those in an unnamed short-term inpatient facility (2 to 3 months treatment duration) for infant and primary

---

36 Reference has been made above [2.2] to Pfeiffer & Strzelecki’s (1990) unqualified recommendations that outcome studies of inpatient psychiatric treatment for children and adolescents employ "powerful statistical techniques and sophisticated designs" (p.852). Reference has also been made to Imrie & Green (1998) providing an uncritical summary [2.31] of the conclusions by Pfeiffer & Strzelecki (1990).
school aged children associated with the Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, Pennsylvania, USA. The children were selected for the study on the basis of (a) their showing externalizing behaviour, and (b) their parents or guardians being unavailable for treatment, due to their being “not in the custody of a parent or relative, if they were likely to be placed outside of the home (e.g., in a foster care) after hospitalization, of if special family circumstances (e.g., excessive distance from the hospital) precluded parent contact with the treatment facility” (p. 77). The authors note that 75% of the children in the psychiatric facility who otherwise met the criteria for inclusion in the study “did not have an available parent who could participate in treatment” (p. 77).

This is in contrast to other units for children this age described in this chapter, where family work was an important component of the treatment. The children were randomly allocated to 45 minute cognitive-behavioral problem-solving skills training sessions (emphasising new ways to solve conflict in interpersonal situations), 45 minute Rogerian counselling sessions (emphasising non-directive, empathic expression of unconditional positive regard) or attendance at 20 minute treatment-contact control sessions, each taking place 2 or 3 times a week for a total of 20 sessions. The outcome variables were derived from the Child Behavior Checklist (Achenbach & Edelbrock, 1983) and the School Behavior Checklist (Miller, 1977b), and administered on 5 occasions: pre-treatment, then 1, 4, 8 and 12 months after completion of treatment.

---

37 Given that 51 out of the 56 children were identified as having as their “primary caretaker” a biological mother (n = 43), step, foster or adoptive mother (n = 6), or other relatives (n = 2), it must be the case that breakdown of the child’s family placement or other “special family circumstances” accounted for the very high proportion of children at this facility who did not have a parent available to participate in treatment.
This study did not meet the criteria for inclusion in the earlier section of this chapter on findings pertaining to the effectiveness of intensive, psycho-social, residential treatment for severely emotionally disturbed children because it represents the results of a treatment study with children selected from an inpatient child unit for the purposes of an experimental study, rather than an evaluation of the effectiveness of that unit. In fact, as mentioned, the facility remains unidentified. There are no positive descriptions provided of its general therapeutic approach or philosophy; the paper notes that when the term “milieu” is used it denotes “a general ward-management program rather that milieu therapy”. It is also stated that “ward activities and structure were directed toward care and management of the children rather than toward specific regimens to treat conduct disorders or other dysfunctions”, although “time-out and seclusion contingencies were included to routinely to manage uncontrollable behaviour” (p. 79).

Nor are the other types of children at the facility (i.e. children not meeting the criteria for inclusion in the study) described. Given the Total Problem Score on the Child Behavior Checklist (CBCL) - T scores of around 78 to 79 reported for the study group at the time of their admission - the included children are clearly showing quite severe behaviour problems. The paper describes only 1½ to 2¼ hours (i.e. two or three 45 minute sessions a week) out of the week for the children in the two treatment groups, and 40 minutes to one hour out of the week for those in the treatment-contact control group. It is difficult to form a clear idea of what takes place in the other 165 hours of the week in this facility, although the absence of mention of professional mental health staff, or the emotional experience of these children kept in a seven day a week facility with no therapeutic involvement with parents or other carers from their life outside the facility, suggests that the brief treatment sessions may have taken place against a backdrop of behavioural control and minimal custodial care.
The authors report that those who received cognitive-behavioral problem-solving skills training sessions showed less problem behaviours at posttest and 12 month followup than those who were in the other two groups. For example, while those attending the Rogerian counselling sessions or the treatment-contact control sessions showed almost no differences on CBCL total problem behaviour score (maximum of 2.5 T score points) between pretest, post test and 12 months followup, the average scores for those receiving cognitive-behavioral problem-solving skills training sessions showed a substantial decrease from pretest (T score = 79) to post-test (T score = 67.5), with the gains being retained at 12 months followup (T score = 68.5). The attrition which took place - attributed to “premature termination of hospitalization”, “child refusal to come to sessions”, and family non-compliance with posttest and followup (p. 79) – was fairly evenly distributed across the three groups at post test (the original 56 subjects were reduced to 47). This is important as it might be expected that the children lost from the study would have poorer outcomes than those who completed both treatment and the requirements of research. No data is provided about the distribution of the 14 cases (25%) lost to 12 months followup, however.

The experimental design (i.e. random allocation to one of three treatment groups) allows some confidence in attributing the positive change to cognitive-behavioral problem-solving skills. While, for example, the average level of total problem behaviour score at followup clearly remains in the clinical range, the lack of change in scores in the other two groups over time strongly suggests that genuine therapeutic gains - that is, changes not attributable to artefacts of instrumentation, attrition and so forth - were achieved through cognitive-behavioral problem-
It needs to be asked, however, whether this study offers strong support for the recommendation of cognitive-based problem-solving skills training packages for children and adolescents undergoing psychiatric hospitalization, as appears in Blanz & Schmidt (2000). While these authors state their conclusions cautiously, their conclusion with regard to effective treatment types in these settings is very specific. As stated, without knowing more about the setting in which the experiment described by Kazdin, Esveldt-Dawson, French & Unis (1987) took place, it is difficult to place the findings into context. It might be asked, for example, whether any signs of developing new conflict resolution skills by children following cognitive-behavioral problem-solving skills training sessions were encouraged and praised in the facility (that is, provided with opportunities for practice and generalisation).

It might also be asked, on the other hand, whether these children found the non-directive Rogerian counselling sessions – intended to provide “the development of a close interpersonal relationship” between therapist and child, and “a corrective emotional experience” (p. 77) – to be very confusing and unhelpful given the context of their placement in a setting where the regard in which they are held can be expected to be highly conditional on their behaviour. It is not surprising that a child between the ages of 5 and 13, in a seven day a week psychiatric facility, whose family placement is breaking down and without a parent or parental figure available to participate in therapy, and whose behaviours demand firm management for reasons of the safety of others, would show little behavioural change following brief sessions of “unconditional positive regard” (p. 79) – conditions unlikely to be met in the facility outside of
the sessions, let alone in their lives outside of the facility. Given the unlikeliness of establishing a genuine relationship under these circumstances, in contrast to the genuine rapport that might be expected to develop between therapist and child working on problem-solving skills relevant to problems in their lives, it is misleading to state that Kazdin, Esveldt-Dawson, French & Unis (1987) compared “problem-solving skills training” to “relationship therapy”, as reflected in the title of their paper.

We return to the question of the comparison of the effectiveness of different treatment types in the following section.
3.5 Summary.

In this section, an attempt is made to summarise the substantive findings [3.51] and the methodological problems [3.52] emerging in this chapter. Following this, the indications for future lines of research are outlined [3.53].

3.51 Substantive findings

An examination of the body of research concerned with the effectiveness of day and residential psychiatric treatment for children and adolescents is not a simple or straightforward task. As argued, much of it suffers from inappropriate data analyses and interpretations, and conclusions are drawn which are not justified by the obtained results. Another major difficulty in approaching this literature is that clinical outcomes are determined according to various procedures or instruments, often unsystematic or non-standardised, which make comparison of the results of one study against another very difficult. While the methodological issues are by no means unique to this area, there is still little evidence of the greater sophistication which has emerged in the general field of psychotherapy and behaviour change research.

As to any question of the general effectiveness of intensive day or residential treatment for severely emotionally and behaviourally disturbed children and adolescents the foregoing review indicated a very broad range, from programs demonstrating quite impressive results sustained at long-term follow-up, on one hand, to programs which appear to have been quite ineffective, on the other.

For example, the evaluation by Zimet, Farley, Silver, Herbert, Robb, Ekanger & Smith (1980) on the effectiveness of the psychoeducational day treatment center at the University of
Colorado Health Sciences Center, Denver, USA with infant and primary school aged children with mainly internalizing or externalizing problems, showed that it had a substantial impact which is sustained at follow-up. Evidence is provided not only of a decrease in problem behaviours relative to pretest, but of problem behaviour levels at followup which is no longer in the clinical range. To take another example, the evaluation by Blackman, Eustace & Chowdhury's (1991) of CASE (Child and Adolescent Services) House, Edmonton, Alberta, Canada, a semi-secure residential treatment program for severely dysfunctional adolescents with serious psychiatric problems, reports a treatment effect size and level of functioning at one to three year follow-up which leaves little doubt about the clinical significance of the outcomes achieved.

On the other hand, the evaluation by Winsberg, Bialer, Kupietz, Botti & Balka (1980) of the residential treatment provided by Kings County Hospital, Brooklyn, New York, USA, to infant and primary school aged children already showing chronic, severe behaviour disorders, failed to show any substantial impact on the problem behaviours.

Other studies show results in between these extremes, where it appears some positive change took place although not enough to bring the youths into the normal range. For example, the evaluation by Kotsopoulos, Walker, Beggs & Jones (1996) of behavioural and academic outcomes for the day treatment program at the Royal Ottawa Hospital, Ontario, Canada, a tertiary care facility for “children with the most severe psychiatric disorders who cannot be managed by other services in the region” (p. 373), suggest that on average a small but significant change in problem behaviours was achieved.
The variation seen in the effectiveness of programs suggests that it is not a sensible question to ask about the general effectiveness of day and residential programs.

With regard to the more interesting question, then, of what child, family or treatment variables account for differences in outcome, there are too few comparable studies to support many general conclusions. While, for example, there is one report (Anderson, Long, Leathers, Denny & Hilliard, 1981) which suggests that the therapeutic preschool they evaluated was more successful with young children showing internalizing behaviour rather than those showing externalizing behaviour [3.21], there is not enough information provided about the preschool to indicate why this might be the case. There is no accumulation of findings to suggest that among children with problem behaviours severe enough to warrant admission to day or residential treatment, one problem type is any more responsive to treatment than the other.

To take another example, there is one small study (Goldfarb, Goldfarb & Pollack, 1966) [3.42, 3.43] which reports that among psychotic infant and primary school aged children in a day and residential treatment centre, those with a "nonorganic" schizophrenic disturbance did best in residential treatment owing to their removal from the putative negative influence of their family. Without comparison against findings from other treatment centres, however, there is no basis for confidence that this interesting finding holds any external validity. That is, an evaluation of any treatment facility is, essentially, a single case-study. Without the accumulation of data from different child units, or from the same unit over an extended period of time, there is no sound basis for inference. The interesting point about this particular study, as noted, is that the finding would not have emerged without the analysis of outcomes being conducted in subgroups based on the clinically meaningful distinction between “organic” and “non-organic” psychosis.
A finding with regard to predictors of outcome which did emerge in more than one study – in fact in three different day treatment programs for infant and primary school aged children (Swan & Wood, 1976; Grizenko & Sayegh, 1990; Grizenko, Papineau & Sayegh, 1993; Kotsopoulos, Walker, Beggs & Jones, 1996) [3.23] - appears to have passed without much comment. There was a finding in the first three studies that while the less able students may show academic improvement, the relatively more able students do not, and may even show deterioration. This appears to be echoed in the finding of the fourth study that while the children did not show academic gains overall during the course of admission, the children weaker in basic receptive language ability showed more gain in reading ability over the course of admission. Stated conversely, the children stronger in basic receptive language ability showed less gain, or even deterioration, in reading ability over the course of admission. Further studies might explore whether this represents, for example, a negative effect of removing from their school emotionally and behaviourally disturbed children who are relatively well functioning academically, or whether the deterioration in academic functioning would take place even if they continued in their school. These findings also suggest that academic deterioration may take place despite behavioural improvement. This might be associated with the chaotic nature of unit schools that include so many children with disruptive behaviours.

With regard to the association between good emotional and behavioural outcomes and specific treatment types, the differences between studies with regard to population, treatment, and method complicate and obstruct the task of examining what convergence there may be in this field. In Chapter 2, the various approaches to summarising results across studies have been described: “tabular asterisks” (which emphasises p values), meta-analysis (which summarise a whole field by a single measure), and the Binomial Effect Size Display (which represents the
treatment effect as differences in outcome rates between the treatment and the control group, expressed as $r$). Their inherent limitations with regard to studying the clinical significance of treatment effects across studies have been outlined. An alternative approach, to examine the raw test scores, was put forward as a way of allowing clinically significant patterns that may be present in treatment effects across studies to emerge. This approach will be illustrated.

There are six studies discussed in this chapter which uses the Total Problem Score of the Child Behaviour Checklist as the outcome variable (Cohen, Bradley & Kolers, 1987; Grizenko & Sayegh, 1990; Grizenko, Papineau & Sayegh, 1993; Grizenko, 1997; Kazdin, Esveldt-Dawson, French & Unis, 1987; Kiser, Millsap, Hickerson, Heston, Nunn, Pruitt & Rohr, 1996) and a seventh which does not report the Total Problem Score but the Internalizing and Externalizing subscales instead (Kotsopoulos, Walker, Beggs, Jones, 1996). The findings of these seven studies, of five different treatment settings, are summarised in the following table (Table 4) and accompanying graph (Fig. 1). It needs to be pointed out that the results presented are standard scores (T scores), as appears in the original reports, rather than raw scores. Although it has the arguable advantage of ease of comparability across studies with different age and gender composition, this introduces unwanted effects of the standardisation process. It is a limitation of the following analysis necessitated by the preference of many authors to express Achenbach rating scores in T-score rather than raw form$^{38}$.

In the following analysis, three treatments will be classified as “less effective”; these are represented towards the left of Figure 1 (Cohen, Bradley & Kolers, 1987; Kiser et al, 1996;
Kotsopoulos et al, 1996). Two treatments which will be classified as “more effective”; these are represented towards the right of Figure 1 (Grizenko & Sayegh, 1990; Grizenko, Papineau & Sayegh, 1993; Grizenko, 1997; Kazdin et al, 1987).

It is also the preference of many authors to perform statistical analyses in this way despite Achenbach’s recommendations in this regard (Achenbach, 1991, p. 148).
<table>
<thead>
<tr>
<th>author</th>
<th>institution</th>
<th>age</th>
<th>problem type</th>
<th>family involvement?</th>
<th>treatment duration (mths)</th>
<th>treatment type</th>
<th>pre treatment M (sd) n</th>
<th>post treatment M (sd) n</th>
<th>6 mths pst-disch M (sd) n</th>
<th>12 mths pst-disch M (sd) n</th>
<th>5 years pst-disch M (sd) n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen, Bradley &amp; Kolers (1987)</td>
<td>therapeutic preschool, Thistletown Regional Centre, Toronto, Canada</td>
<td>pre-school</td>
<td>internalizing &amp; externalizing subgroup</td>
<td>participation in preschool activities, &quot;regular contact&quot; with social worker</td>
<td>12 to 24</td>
<td>day admission, specialist educ. &amp; health profs.</td>
<td>67.60 (9.2) n=9</td>
<td>61.08 (9.0) n=9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grizenko &amp; Sayegh (1990)</td>
<td>Lyall Preadolescent Day Treatment Program, Douglas Hospital, Verdun, Quebec, Canada</td>
<td>infant &amp; primary</td>
<td>internalizing &amp; externalizing</td>
<td>family therapy</td>
<td>7</td>
<td>day admission, psychodynamic</td>
<td>78.7 (6.6) n=23</td>
<td>54.1 (11.2) n=23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grizenko, Papineau &amp; Sayegh (1993)</td>
<td>as above</td>
<td>as above</td>
<td>as above</td>
<td>as above</td>
<td>4.4</td>
<td>day admission, psychodynamic</td>
<td>82 (4.6) n=15</td>
<td>60 (8.4) n=15</td>
<td>58 (10.40) n=14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grizenko (1997)</td>
<td>as above</td>
<td>as above</td>
<td>as above</td>
<td>as above</td>
<td>6.1</td>
<td>day admission, psychodynamic</td>
<td>81.7 (6.4) n=33</td>
<td>56.3 (7.5) n=33</td>
<td>67.6 (8.7) n=33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kazakhstan et al (1987)</td>
<td>unidentified inpatient facility, Pennsylvania, USA</td>
<td>infant &amp; primary</td>
<td>externalizing subgroup</td>
<td>none</td>
<td>2 to 3</td>
<td>individual cognitive-behav. problem solving training, in inpatient facility</td>
<td>79 (5.5) n=20</td>
<td>67.5 (9.8) n=18</td>
<td>68.5 (8.4) n= (na)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kiser et al (1996)</td>
<td>two unidentified day treatment facilities, Tennessee, USA</td>
<td>infant, primary &amp; adol</td>
<td>internalizing &amp; externalizing</td>
<td>family therapy</td>
<td>3.9</td>
<td>day admission</td>
<td>70.08 (na) n =114</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kotso-poulos et al. (1996)</td>
<td>day treatment program, Royal Ottawa Hospital, Ontario, Canada</td>
<td>infant &amp; primary</td>
<td>mainly externalizing</td>
<td>family counselling and parent training by a child &amp; youth counsellor</td>
<td>18</td>
<td>day admission, &quot;multimodal&quot;, behavioural, cognitive-behavioral, medication</td>
<td>INT=68.9 (10.6) n = 46</td>
<td>EXT=73.5 (8.7) n = 46</td>
<td>INT=63.8 (14.8) n = 46</td>
<td>EXT=66.3 (14.7) n = 46</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Evaluative studies reporting CBCL Total Problem Scores (except Internalizing and Externalizing scores where indicated)
Figure 1. Graphic summary of evaluative studies reporting CBCL Total Problem Scores (except Internalizing and Externalizing scores where indicated)
The graphic display of average scores highlight the following differences leading to the classification into the “less effective” and the “more effective” treatments.

1. The pre-treatment level of problem behaviours for the three less effective treatments is much less pronounced than for the more effective treatments. The pre-treatment Total Problem Scores (or in the case of Kotsopoulos et al, 1996, the Internalizing and Externalizing Scores) of the three less effective treatments are below or not far above a T score of 70 (97.7th percentile). On the other hand, the pre-treatment Total Problem Scores of the two more effective treatments are above a T score of 75 or even 80.

2. The amount of change from pre-treatment to post-treatment is smaller in T score units in the less effective treatments than in the more effective treatments. (Given the way in which T scores are derived, the lower pre-treatment scores of the less effective treatments means that there would be a even larger difference between the three less effective and the two more effective treatments with regard to the amount of change expressed in raw units).

3. For the less effective treatments, the post-treatment Total Problem Score (or in the case of Kotsopoulos et al, 1996, the Internalizing and Externalizing Scores), remains quite high. The post-treatment level of Total Problem score (or Internalizing and Externalizing Scores) is above 60 (borderline clinical range) or even 63 (clinical range) for the less effective treatments. For the more effective treatments, there is some variation in the post-treatment Total Problem Score, to be discussed below.

Focusing now on the results obtained by the two more effective treatments, there are the following differences between the treatment effects reported by Grizenko and others at the Lyall
Preadolescent Day Treatment Program of the Douglas Hospital, Verdun, Quebec, Canada and the treatment effect demonstrated in the study by Kazdin, Esveldt-Dawson, French & Unis (1987).

1. The pre-treatment level of problem behaviours for the treatment evaluated by Grizenko and others are more pronounced, or equal, to that of the treatment evaluated by Kazdin et al (1987).

2. The amount of change from pre-treatment to post-treatment is larger (in T score units) for the Grizenko studies than the treatment evaluated by Kazdin et al (1987). (The difference in the amount of change expressed in raw score units will be even larger).

3. The treatment evaluated by Grizenko and others results in post-treatment problem scores in the normal range (i.e. less than 60), sustained at 6 months post-discharge. The post-treatment problem score in the treatment evaluated by Kazdin, Esveldt-Dawson, French & Unis (1987) remains well in the “clinical range” both immediately post-treatment and 12 months post-discharge.

On the basis of the available evidence, and using standardised scores as the unit of measure of treatment effect, the superiority of the treatment evaluated by Grizenko and others over the treatment evaluated by Kazdin, Esveldt-Dawson, French & Unis (1987) is quite apparent. In fact the amount of change reported by Grizenko (1997) from pre-treatment scores to 5 year followup, and the level of post-treatment problem behaviours at that time, is quite similar to that seen at 12 months post-discharge by Kazdin et al (1987).
The present results are not explainable by different levels of attrition. The children in the treatment group in the Grizenko & Sayegh (1990) study are described as 23 consecutive admissions, with no attrition. The children in the treatment group in the Grizenko, Papineau & Sayegh (1993) study appear to be 15 consecutive admissions, again with no attrition$^{39}$. On the other hand, the 20 children commencing in the treatment evaluated by Kazdin et al (1987) was reduced to 17 at post test and reduced further$^{40}$ by 12 months follow-up. It would be expected that those who drop out of treatment studies are likely to have a poorer outcome than those who remain. Therefore any artifactual inflation of therapeutic results through attrition is likely to favour the treatment evaluated by Kazdin et al (1987) over the treatment evaluated by Grizenko et al.

These seven studies in five different settings do not allow conclusions to be made with confidence about the treatment variables associated with good outcomes. Yet, as seen above, Blanz & Schmidt (2000)’s review conclude “cautiously” (p. 703) on the basis of only one study - Kazdin, Esveldt-Dawson, French & Unis (1987) - that inpatient treatment for children is “often beneficial” when it includes special aspects such as “a cognitive-based problem-solving skills training package”. On the basis of the limited evidence available, a review which included both day and inpatient treatment might also cautiously conclude – on the basis of the series of studies by Grizenko and others - that multimodal day treatment can be beneficial when it is conducted along “psychodynamic” lines. It might equally be argued along these lines that that a multimodal day treatment based on behavioural management of problem behaviours and

$^{39}$ The children in the Grizenko (1997) study were 33 children described as being from the same cohort as the two earlier studies, “80% of the children who were sequentially admitted and who completed the day treatment program between 1988 and 1990”. As the N of 33 exceeds 80% of the combined N of the two earlier studies, it may be the case that the two earlier studies were subsets of a cohort of approximately 41 admissions.

$^{40}$ As mentioned, while the attrition is provided for each group at posttest, it remains unspecified at 12 months followup.
cognitive-behavioural skills training, combined with the broad use of stimulant medication (Kotsopoules, Walker, Beggs, Jones, 1996), is not very effective.

The present illustration leads to further interesting questions and observations. It might be asked what is the difference between the individual sessions of “cognitive–behavioural problem-solving skills training (PSST)” that was shown to be so effective by Kazdin et al (1987) and the “ongoing, intensive social skills training within the class” (p. 374) that appears to have been much less potent in the treatment evaluated by Kotsopulos et al (1996). While both studies make reference to Kendall & Braswell (1985, 1993) as offering a model for the work, it is made clear in Kazdin et al (1987) that efforts were made to modify this general approach to:

“emphasise interpersonal situations with significant others (e.g. parents, siblings, teachers, peers) and to include opportunities for individualizing content and addressing referral concerns and situations in which the child had engaged in antisocial and oppositional behaviour. The treatment combines cognitive and behavioural techniques to teach problem – solving skills (e.g., generating alternative solutions, means-ends and consequential thinking, and taking the perspective of others) that the child can use to manage interpersonal situations” (p.78).

On the other hand, it is difficult to gauge from the brief mention in Kotsopulos et al (1996, p. 374) how the cognitive-behavioural strategies were implemented, especially as it reported that it took place in a class setting. If the implementation of the basic cognitive-behavioural model was very different in these two settings, this suggests that the “branding” of the clinical model indicates very little about whether a genuinely therapeutic process is taking place; that is, the
claim that a particular model of therapy is being implemented may tell us nothing about its effectiveness.

Furthermore, when a treatment is demonstrably effective, the identified brand of treatment may tell us nothing about the active ingredients. To illustrate, it is very striking that the modifications in the general cognitive-behavioural model which took place in the treatment evaluated by Kazdin et al (1987) appears to have resulted in a method similar in important ways to the “Life Space Interview” developed by Redl (Redl & Wineman, 1951, 1952), albeit unacknowledged. The latter, an active treatment method for intervening directly whilst a young person is in a state of emotional crisis and negative behaviours, developed from psychoanalytic principles, was very influential in the 1950s in the USA and also in Australia. (It formed, in fact, the basis of the milieu treatment at Arndell Child and Adolescent Unit – where the present study described in Chapters 4, 5 and 6 was conducted – until the mid-1990s).

Similarly, it might also be asked whether the type of psychodynamic day treatment program evaluated by Grizenko and others is able to be implemented effectively and with similar effectiveness in other settings. There also remains other interesting, related, questions of whether effective programs sustain their effectiveness over time, whether programs differ in ways not readily apparent from published reports that may still significantly impact upon its effectiveness (eg. level of morale, charismatic leadership, theoretical cohesion, and so forth).

The following conclusions are made on the basis of the present review.

1. There is no reliable empirical basis to suggest that any specific treatment is more effective than any other in the day or residential treatment of severely emotionally disturbed children and adolescents.
2. More particularly, the current popular view, as appears in the review by Blanz & Schmidt (2000), that there is evidence to recommend cognitive-behavioural treatment within such settings appears to rest largely on one study in one setting (Kazdin, Esveldt-Dawson, French & Unis, 1987). An illustrative examination of that study in the context of seven studies of five treatment programs utilising comparable outcome measures shows that there is an equally thin basis to argue, on the basis of findings from a series of studies in a very different setting (Grizenko & Sayegh, 1990; Grizenko, Papineau & Sayegh, 1993; Grizenko, 1997), that there is a basis to prefer comprehensive psychodynamic treatment.

3. Furthermore, this illustrative examination suggests that there is no reliable association between the specific treatment type with which a program identifies itself and its effectiveness.

4. Finally, a closer examination of the treatment evaluated by Kazdin, Esveldt-Dawson, French & Unis (1987) suggests that a sophisticated cognitive-behavioural, problem-solving training program for emotionally and behaviourally disturbed young people which focuses on interpersonal issues may not be readily distinguishable from the “Life Space Interview” (Redl & Wineman, 1951, 1952) - an intervention method based on psychoanalytic principles developed for the same population in the 1940s and 1950s.

These conclusions suggest that, contrary to the assumptions of the “empirically validated” treatment movement, it is not at all clear what role, if any, clinical orientation – defined as the specific treatment type which a treatment program is said to represent - plays in its effectiveness.
3.52 Methodological problems

The common methodological problems found in the studies reviewed in this chapter are briefly summarised under the following headings: research design; data analysis; variables and instruments.

**Research design:**

1. Some follow-up studies lacked sufficient structure, such as pre-treatment data, to possess even minimal internal validity. This was usually related to the fact that they were not planned in advance (e.g. Eisenberg, Lansdowne, Wilner & Imber, 1962) [3.21].

2. With regard to control groups, many of the studies addressing the effectiveness of institutional treatments lacked adequate control groups even when the research question and the data analytic method of choice requires one (e.g. Grizenko & Sayegh, 1990) [3.23]. There are also studies using a control group systematically biased *against* demonstrating improvement in the treatment group (e.g. Mitchell, Rothwell & Burtenshaw, 1975) [3.21] as well as studies using a control group systematically biased *towards* demonstrating improvement in the treatment group (e.g. Woollacott, Graham & Stevenson, 1978) [3.21].

3. Another frequent limitation was post-treatment data not being collected beyond discharge. As argued, not only does this provide no information with regard to the durability of any clinical change, post-tests conducted at the point of discharge is vulnerable to the challenge that this may be capitalizing on “random variation” in the child's symptomatology [3.21]. It is of interest, however, that on two well-designed
studies which included long-term follow-up (Zimet et al., 1980; Blackman, Eustace & Chowdhury, 1991) [3.51], the clinical gains were maintained or increased.

Data analysis:

4. The central data analytic problem was the common misrepresentation of $p$ values as an expression of effect size (discussed in 2.41). Effect size was often not reported, or presented in an uncertain and unsystematic manner. Another important aspect of outcome in considering the clinical significance of the results, the child’s post-treatment level of functioning, was also often neglected.

5. The data analyses usually assumes a homogeneity in treatment response\(^4\). An exception, the Goldfarb, Goldfarb & Pollack (1966) study, to which numerous references have been made throughout this chapter [3.42, 3.43, 3.51], shows how the analysis of outcome data in clinically meaningful subgroups can provide results which would not otherwise emerge.

6. The often very small numbers presented a limitation to data analysis. Multiple regression analyses and multivariate techniques to test for predictors of outcome were often performed where the numbers were not adequate (e.g. Grizenko, Papineau & Sayegh, 1993) [3.23].

7. The issue of drop-outs from the treatment programme or from follow-up, or the selection of study subjects from a clinical population on the basis of the availability of complete

\(^4\) This is one of the “myths of psychotherapy research” discussed by Sanford (1953), Colby (1964) and Kiesler (1966), among others.
data sets, was often not adequately addressed in the effectiveness studies. It very likely resulted in artifactual inflation of the measured effectiveness of the treatment under evaluation (eg. Grizenko (1997) [3.23].

Variables and instruments

8. The methods for generating data were often vulnerable to bias and the interests of the researchers. Methods likely to produce data of doubtful worth included (a) unstructured follow-up interviews; (b) depending on the mother's report rather than testing the child directly on variables such as language ability; and (c) instruments developed specially for a study with no demonstrated reliability and validity or other psychometric properties. Even where standardised tests were utilised, there was often insufficient care to protect the integrity of the process from bias associated with the special interest, expectations or wishes of those participating in the data generation.

9. The variables and processes under examination are often those relatively easy to capture but of limited theoretical interest. For example, while most day treatment programmes which have been evaluated, especially those for younger age groups, emphasised the importance of family treatment or family involvement, the influence of family functioning on outcomes is a topic which remains largely ignored.
3.6 Conclusions with regard to future lines of inquiry in the area.

The foregoing considerations lead to the following conclusions with regard to the design of an evaluative research program which is likely to be viable and robust, on one hand, while producing data of clinical and theoretical interest, on the other.

3.6.1 Pragmatic constraints on research question and design

The methodological difficulties in conducting evaluations of intensive psychosocial day and residential treatment for severely emotionally disturbed children and adolescents, as outlined, have ready solutions in principal. It is important, however, in planning a viable evaluative research program that the pragmatic constraints need to be recognised from the outset.

The most intractable among them is likely to be the difficulties in the recruitment of a non-treatment control group. There are the obvious constraints on randomly assigning a pool of severely emotionally disturbed children whose placement at home and or their school has broken down, to either a treatment or a non-treatment control group, and then ensuring that the families of those assigned to the control group do not receive treatment elsewhere for a period equivalent to treatment and follow-up. The role of public child and adolescent mental health services and non-government organisations usually requires that those referred receive a response. That is, these agencies are required to respond to the behaviour of children whose placements with their families and schools are breaking down, yet remain outside of the juvenile justice system, irregardless of the effectiveness of its treatments and interventions.

A quasi-experimental design, a comparison between two “naturally occurring” treatment groups, avoids this constraint. It might be suggested, for example, that admission to a minimum-care
facility (such as the child-adolescent unit at a USA state hospital evaluated by Treffert, 1969) represents a "naturally occurring" non-treatment control group available for comparison against a more sophisticated and intensive treatment unit. However, such a design will still face a pragmatic problem intrinsic to comparison studies. Even if institutional resistances to participation in such a comparative study could be overcome, the problem of competitive interests biasing the data obtainable from children, their families and the staff needs to be addressed. Some of these influences, perhaps the most pernicious to the integrity of the study, are likely to operate outside the explicit, readily acknowledged operational parameters of such a study. These neglected issues were discussed in the section on pragmatic issues in Chapter 2 [2.3].

To provide just one example touching on the issue of attrition, it is a small private matter remaining “under the radar” of the research protocol when a researcher shows decreased ardour in encouraging the persistence of a research assistant attempting to obtain follow-up data from a child and family when evidence of a poor outcome is already at hand. The obvious in-principal solution of disinterested researchers located outside of the treatment centres and maintaining “blindness” to treatment condition introduces further pragmatic constraints. The unexpectedly strong influence of assessor expectations on testee responses has been discussed in detail in relations [2.52].

The alternative, as demonstrated by Kazdin, Esveldt-Dawson, French & Unis (1987), of conducting random controlled trials of distinct individual treatment types within an actual child facility will enjoy the clear advantage of experimental design with regard to internal validity. However, the dominant treatment philosophy in the host facility for any such study can be expected to form an important background to the specific treatment types of interest, with the
interaction between background and foreground likely to have a powerful influence on the apparent effectiveness of the latter. Clinicians with experience in treatment facilities or institutions are aware of the many different ways in which their individual work with patients can be strongly augmented or undermined by the dominant, implicit, “background” philosophy of the facility, and the special interests which may be active within it. In this sense, studies of this type focus only on a small part of the therapeutic process in a setting; in the case of Kazdin, Esveldt-Dawson, French & Unis, 1987, the “foreground” of 15 hours of individual treatment which takes place during the two to three months of admission and not the “background” of the other 1300 to 2100 hours of the admission.42

Furthermore, such a paradigm is not suitable for more intensive, broad and sophisticated treatments as they are unlikely to be deliverable in this way. This is similar to the limitations of the single-case experimental design discussed in Chapter 2 [2.42]. For example, in the study by Kazdin, Esveldt-Dawson, French & Unis (1987), while the superiority of the cognitive-behavioural problem-solving skills training sessions over the control group seems impressive, it seems very unlikely that the Rogerian counselling sessions could have been conducted coherently and with integrity within the overall philosophy of the facility as described.

In contrast, naturalistic studies which address the question of what child, family and treatment variables or processes distinguish those who show good outcome from those who show poor outcome in a setting is much less likely to arouse institutional resistance, and therefore the associated threats to the integrity of the research.

42 Calculated from the information available in Kazdin, Esveldt-Dawson, French & Unis (1987).
In this way, an evaluative research program which commences with a realistic appraisal of pragmatic constraints of the clinical setting of interest leads to the selection of a naturalistic research design, and the selection of the research question: *what variables account for individual differences in outcome?* Naturalistic studies of what actually takes place in child units can proceed while having minimum impact on the clinical processes themselves. For these same reasons, it is concluded that a study which addresses the question *how effective is the treatment under examination?* is not likely to be viable in an actual clinical setting.

It is argued that the question at the centre of the much smaller number of studies reviewed in 3.4 can be adequately addressed. On the other hand, it is argued that - contrary to previous reviewers who view the application of particular research methods such as control groups, powerful statistical techniques and so forth as the way forward in this area [2.2] - the research question at the centre of the much more numerous studies reviewed in 3.2 and 3.3 cannot be adequately addressed. It is unlikely that the pragmatic difficulties will be overcome.

The other methodological advantage of a research design of the former type is that the unavoidable problems, of bias associated with (a) the selection of the study sample from the clinical population, and (b) natural attrition, will be less crucial. While missing data can be expected to artifactually inflate outcomes for any treatment program, there is no similarly compelling reason to expect that missing data is likely to change the observed *relationship* between predictors and outcome.
3.62 The variables to be studied

As for which predictor variables might be studied, very basic issues such as the influence on outcome of the child's age, gender and problem type, or family functioning, structure and socio-economic status, or treatment length and type, remain to be examined. While it lacks internal validity of an experimental design, a naturalistic study has the potential to study important events in the clinical setting which are unlikely to be able to be brought under experimental control.

It also offers the possibility of a “natural experiment”. As it turned out, the dominant treatment type changed abruptly during the decade that evaluative research was taking place at Arndell Child and Adolescent Unit, from where the data for the study described in the following chapters (4, 5, 6 & 7) was drawn. This meant that, entirely fortuitously, treatment type was able to be included among the predictors of outcome.

3.63 Model of data analysis

As discussed in Chapter 2 [2.4], there is an inevitable tension in data analysis between retaining the clinical significance which pertains to datum in the context of the individual case, on one hand, and making useful generalisations about a population or a variable which transcend the particulars of the individual case, on the other. While the assumption of homogeneity of patient response to treatment underlying the group model of data analysis (Kiesler, 1966), discussed above [2.42], needs to be rejected, the problem of drawing valid and clinically meaningful generalisations from data extracted from a large set of unique individuals remains to be addressed.
The clinical importance of distinguishing different internal processes underlying similar external behaviours was illustrated, for example, by Megargee’s (1966) distinction between undercontrolled and overcontrolled personality types in youths incarcerated for acts of extreme violence [2.41]. In a similar way, the value of conducting analyses within clinically meaningful subgroups was demonstrated in the small study by Goldfarb, Goldfarb & Pollack (1966) [3.52].

One way, then, to draw general conclusions from an analysis of predictors of outcome which is both statistically robust and clinically meaningful may be to conduct the analyses within clinically meaningful subgroups. The unsophisticated application of statistical procedures in predictor studies and the confusions evident in the conclusions drawn from them has been discussed [2.51]. Also discussed was the inclusion of pre-treatment pathology among predictors of post-treatment pathology in multiple regression analyses leading only to the unremarkable finding that the best predictor of post-treatment functioning is pre-treatment functioning. A multiple linear regression analysis taking place within clinically meaningful subgroups allows for the differential role of predictors to outcome to emerge from the heterogeneity of the subjects.

As to what particular clinical subgrouping – that is, in what ways can the clinical population of interest be subdivided to bring out the different patterns of predictors operating within them? – it has been argued [2.52] that it may be more productive to consider differences between a child’s presentation at home (as rated by parents) and at school (as rated by teachers) as meaningful data rather than measurement error. This has particular relevance to the treatment of interest. Intensive psychosocial day and residential treatment for severely emotionally disturbed children and adolescents offer mental health and special education professionals working together in the
one setting. It is distinguished in this way from less intensive treatments, such as outpatient child and family mental health treatment or placement in a special class within a school setting. By virtue of its offering a service for youths showing difficulties in both domains, these units offer a rare opportunity to conduct evaluative research which addresses not just the severity of youths’ difficulties, but also the pervasiveness of their difficulties.

The careless conclusions appearing in influential publications about the impact of intensive psychosocial day and residential treatment on severely emotionally disturbed children and adolescents has been described [2.51]. Concern has been expressed about the impact of these conclusions, which may be to direct treatment away from the most disturbed children from the most disadvantaged family situations. The study reported in the following chapters focuses specifically on this patient group, which – it has been argued – has been ill-served by researchers and reviewers of the research. This study is specifically designed to allow differences, if any, to emerge between predictors of outcome for the most severely disturbed children – those showing severe problems both at home and at school – and predictors of outcome for children showing more moderate, less pervasive difficulties.
CHAPTER 4.

METHOD

4.1 Introduction to this Chapter

This chapter describes the method of a naturalistic study of the predictors of outcome for 159 severely emotionally disturbed children and adolescents admitted to intensive psycho-social treatment at Arndell Child and Adolescent Unit, a day and residential treatment unit in Sydney, Australia between 1990 and 1999.

The research question was: what child, family and treatment variables predict outcome for the children and adolescents treated at Arndell Child and Adolescent Unit from 1990 to 1999?
4.2 The population under study

Arndell Child and Adolescent Unit was a mental health/education Unit located at North Ryde, NSW, jointly administered by the Department of Child & Adolescent Psychiatry of Royal North Shore Hospital and the NSW Department of Education. It provided over the period of this study multi-modal treatment (individual, group and family treatments) to children and adolescents between the ages of 3 to 18 and their families. The Junior Program offered outpatient and day treatment to children aged between 3 and 7, the Middle Program offered outpatient, day and residential treatment to children between the ages of 8 to 13, and the Senior Program offered outpatient and day treatment to adolescents aged between 13 and 18; this roughly corresponded in age to infants, primary and high school.

From the beginnings of the Children’s Unit on the grounds of North Ryde Psychiatric Hospital in 1958 - when it was the only specialised psychiatric Unit for children in the southern hemisphere, at that time receiving referrals from as far afield as the British Commonwealth countries of India and New Zealand - through the commencement of “Arndell” on a new site in 1979 (Wotton & Laganza, 1981), the close and active collaboration between educational and health staff in one “unit” was reflected in the interchangeable reference to either the Arndell “Program” or Arndell “School”.

Groups of mental health staff were allocated to a particular class (or classes) of students with its associated teacher and teacher aide, so that in addition to consistent educational staff each child was allocated a case-manager from among the clinicians (a psychologist, psychiatrist, social worker, or occupational therapist) as well as a primary and secondary nurse. These groupings were identified as “Teams” and the intense nature of the work meant that the primary
identification of a staff member was with his or her “Team” and the children under their care, rather than their discipline.

Children and adolescents referred to Arndell typically showed severe disturbance, e.g. problem behaviours, poor modulation of emotional expression, inadequate social and personal skills, and poor peer relations (MacDonald, 1991). Often their school placement was at risk or had already broken down, and the families typically presented in a state of crisis.

While its role and its programs changed over time, it remained until the mid-90s the only state inpatient psychiatric facility for primary school aged children in NSW and an important tertiary treatment and consultation centre for metropolitan and regional NSW. The changes which took place at that time included the reduction of inpatient treatment and an associated reorientation towards increased outpatient treatment, a reduction in focus from a state-wide service to the needs of the Sydney metropolitan health area in which it was located, and the rejection of the psychodynamic milieu-therapy - which has been the dominant treatment philosophy since the 1960s - in favour of a more generic style of child psychiatry incorporating, as described in Chapter 1, “empirically-based” cognitive-behavioural and behavioural therapy. These changes marked an abrupt transition in the way services were provided in NSW for its most severely emotionally and behaviourally disturbed children, leading eventually to its name change in the late 90s to “Coral Tree Family Unit”. The separation of the education and mental health staff into two operationally and administratively independent halves which also took place at this time is represented by the school retaining the name of Arndell SSP (School for Special Purposes) while the mental health component marked a deliberate discontinuity from the traditions and often turbulent history of Arndell through its adoption of a new name.
The systematic collection of data at Arndell was introduced in July 1990 and ended in October 2001. Its introduction was staged over a 6 month period, commencing in the Middle Program, then the Junior and Senior Programs in early 1991. Approval from the Royal North Shore Hospital Medical Research Ethics Committee was sought and granted at various stages of the project, in keeping with the additions and changes to the research protocol and the requirements of external agencies (e.g. RADGAC) from whom additional research funding was sought [Appendices A & B].

Parents were approached for participation in the evaluative research program in the following way. Parents were presented with informed consent forms with regard to their participation in the study at registration, on their first attendance at the Unit. The form was designed in such a way that they could agree to participate in the study, yet disagree to have their child’s school approached [Appendix F]. If they agreed to participate, and their child subsequently underwent inpatient or daypatient treatment, parents were presented with another informed consent form at discharge to seek consent to follow-up [Appendix G].

For the children of parents who consented, and who were admitted to inpatient or day treatment, their teachers were also contacted and asked to provide information. This contact took place only after the child was accepted into treatment to avoid potential bias in the teacher’s reports arising from their interest in having the child accepted at Arndell.
4.3 Variables and measures

The predictor variables included in the study will be described in three blocks: child variables, family variables, and treatment variables. Following this, the outcome variables will be described.

4.3.1 Child variables

There were six child variables included as predictors of outcome in the study. Two were socio-demographic characteristics and four were clinical variables. They were collected at the time of the child’s registration at the Unit or during their admission.

The two socio-demographic variables were gender and age at registration.

The clinical variables were the severity of problem behaviours reported by parents and severity of problem behaviours reported by teachers, and two different aspects of their clinical presentation, externalizing versus internalizing behaviours and psychotic versus nonpsychotic.

The severity of problem behaviours reported by parents was measured by the Child Behaviour Checklist. The Achenbach scales in their parent and teacher forms were discussed in detail in Chapter 2. As can be seen on Table 5, the indices of test-retest reliability are quite sound, but the level of inter-rater agreement is less so.\footnote{The Pearson correlation coefficient may underestimate disagreement if there are systematic differences between mothers and fathers (for example, if fathers report less problem behaviours than mothers because they have lesser exposure to their children, and so forth). The intraclass correlation would be more informative in this regard (Bartko, 1966; Bartko & Carpenter, 1976).} This is unsurprising given the nature of these scales, as discussed \cite{2.52}, and suggests that the reliable use of this instrument to measure change over time requires the rater to be held constant.
The same issue arises with the Achenbach scale in the teacher form – the Teacher’s Report Form - from which a measure of severity of problem behaviours reported by teachers was obtained (Table 6). Given that the reported inter-rater agreement was obtained from pairs of teachers who “saw the pupil in different classes or under conditions that differed in other ways” (Achenbach 1991b, p.60), it is difficult to know the sources of their disagreement. However, the test-retest reliability is again encouraging and supports the repeated use of the instrument with the rater held constant.

<table>
<thead>
<tr>
<th></th>
<th>test-retest reliability</th>
<th>inter-rater agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(one week)</td>
<td>(mothers and fathers)</td>
</tr>
<tr>
<td>Total Problems</td>
<td>.93</td>
<td>.76</td>
</tr>
</tbody>
</table>

**Table 5.** Reliability of the Child Behavior Checklist (Pearson correlation coefficient) (from Achenbach 1991a)

<table>
<thead>
<tr>
<th></th>
<th>Test-retest reliability</th>
<th>Inter-rater agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(15 days)</td>
<td>(teachers)</td>
</tr>
<tr>
<td>Total Problems</td>
<td>.95</td>
<td>.60</td>
</tr>
</tbody>
</table>

**Table 6.** Reliability of the Teacher's Report Form (Pearson correlation coefficient) (from Achenbach 1991b)

As mentioned [4.21], care was taken with regard to the potential vulnerability of teacher ratings to external demands. A teacher with a very difficult child being assessed for admission to the Unit might be expected to be motivated to more readily report problem behaviours under the expectation that this might increase the chances of the child receiving appropriate help (and out of their classroom!). It is for this reason that the teacher at the school the child attended
immediately prior to admission to the Unit was asked to complete a TRF only after the child’s transfer took place\textsuperscript{44}.

With regard to the child’s clinical presentation, as outlined above [3.2], the dimensions of internalizing and externalizing behaviours are widely considered to be clinically meaningful and have been identified in factor analytic studies of the psychopathological symptoms of childhood (Achenbach, 1985; Quay, 1986). While internalising and externalising scores are readily derived from the CBCL, the present study aimed to avoid the types of methodological weaknesses common to the use of these scores in previous outcome studies.

First, there is often confusion between, on one hand, regarding internalizing and externalizing as pertaining to a child’s diagnostic category, and, on the other, regarding these terms as denoting \textit{behaviours} which are exhibited to various degrees by all children. In the absence of standard methods for categorising children as internalizing or externalizing, most empirical work involves the latter approach. However, findings about changes in externalizing or internalizing scores in a group are often mistakenly reported as if they were findings pertaining to types of children. It is one thing, for example, to conclude that a program lead to a decrease in the level of externalizing behaviour, and it would be quite another to conclude that externalizing youths showed the most change. Returning to the earlier example of the underinhibited and overinhibited subtypes of youths incarcerated for extreme assaults (Megargee, 1966), a punitive regimen may well achieve reductions in the average level of externalizing behaviour in a milieu but this may be due to decreasing externalizing behaviour among the overinhibited group.

\textsuperscript{44} In the case of students in high school, the principal of the previous school was asked to have the TRF completed by the teacher who knew the student the best. This was often the year co-ordinator or school counsellor.
Second, another common methodological weakness in the application of multiple regression techniques using the Achenbach scales is the inclusion of Internalizing and Externalizing scores as predictor variables of outcomes, while the outcomes are also based on post-treatment Achenbach scores. Pre-treatment Internalizing and Externalizing scores, acting as proxies for pre-treatment pathology, lead to the unremarkable finding that the best predictor of post-treatment pathology is pre-treatment pathology, as discussed [2.51].

In the present study, there was interest in internalizing and externalizing as characteristics of the child’s pre-treatment clinical presentation. A method for classifying children with regards to these dimensions needed to be developed as many of the children referred to the Unit were characteristically showing both severe internalising and externalising problems, and, as mentioned, a predictor variable needed to be developed which was not a proxy for pre-treatment morbidity. For these reasons, the following method was adopted: parent internalizing and externalising scores were converted into an externalizing/internalizing index (e/i index), which would reflect the relative prominence of externalizing versus internalizing problems.

\[
\text{e/i index} = \frac{\text{raw externalizing score}}{20} - \frac{\text{raw internalizing score}}{13} + 2.5
\]

The rationale for the index is as follows.

First, as the raw externalizing score necessary to reach the “clinical” threshold (Achenbach, 1991a) is higher than the raw internalizing score necessary to reach the “clinical” threshold, each was standardised as a ratio of the “clinical” score. The clinical cut off scores for boys and girls, in the age ranges 5 – 11 and 12 -18, on the CBCL appear in Table 7 below. The clinical

45 This ruled out Achenbach’s (1991) own suggested methods (p. 66) which involves cut-offs.
scores incorporated into the index was based on the boys scores as boys were in the clear majority in the Unit. No separate index was formulated for the girls as it was an aim of the present study to deal in raw scores so that gender differences, as a predictor of outcome, could be addressed directly (i.e. not pre-adjusted to deviation from gender–based community norms).

The use of raw scores rather than T scores in research use of the Achenbach scales is consistent with Achenbach’s own recommendations, as discussed [3.51], and is especially pertinent in the present application in view of the arbitrary methods for assigning T scores to extremely high raw scores owing to their rarity in the normative population. (Achenbach, 1991, p.189)

<table>
<thead>
<tr>
<th></th>
<th>Internalizing</th>
<th>Externalizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>boys (4-11)</td>
<td>13</td>
<td>20</td>
</tr>
<tr>
<td>boys (12-18)</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>girls (4-11)</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>girls (12-18)</td>
<td>18</td>
<td>16</td>
</tr>
</tbody>
</table>

**Table 7.** Clinical cut-offs for externalizing and internalizing scores on the CBCL. (from Achenbach, 1991)

Second, the constant of 2.5 was added to bring all values of the e/i index into the positive range, then divided by 5 so that they fall in the range .1 to .80, for convenience.

A lower index, then, represents a predominantly internalizing presentation and a higher index representing a predominantly externalizing presentation. No meaning is imputed to the absolute numerical value of the e/i index in any individual case. Its construction served merely to
quantify for each case the relative prominence of externalizing versus internalizing behaviours in their presentation independent of the level of disturbance.

As for the diagnosis of psychosis, while the CBCL is well-suited to categorise problem behaviours into internalizing and externalizing types, its empirically-derived, dimensional approach is not very sensitive to this important aspect of the presentation of children. In contrast, psychosis is given very prominent attention in the conventional, discrete groupings of psychiatric diagnosis. This is likely attributable to the very small presence of psychotic behaviours in empirical surveys owing to its rarity in childhood and early adolescence. On the Achenbach scales, items pertaining to psychotic disturbance of thought appear together with items pertaining to obsessionality in the syndrome scale of “Thought Problems” 46, and are not a suitable method for the diagnosis of psychosis. Whether psychosis is regarded as a distinctly organic condition, as in the current psychiatric view, or as representing a family communicational disturbance, as in the influential family therapy literature of the 1950s and 1960s (Hahlweg & Goldstein, 1987), children diagnosed as psychotic attracted a response in the Unit quite different from the children with the more common internalising and externalising behaviours. Even in the milieu-therapy era – the first half of the period covered in this study - when a psychodynamic approach to the child and to his family was dominant, these children tended to be referred to medical staff and to receive medication. This was in contrast to the treatment philosophy of that time which was to gradually reduce and often discontinue medication as the containing effect of the milieu took effect. For these reasons, it was felt important that the variable of psychotic versus non-psychotic diagnosis was included,

46 In addition to the Total Problem Score and the broader factor scores of Internalizing and Externalizing examined in the present study, the conventional scoring of the CBCL and the TRF also produces eight syndrome scores: Withdrawn, Somatic Complaints, Anxious/Depressed, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, Aggressive Behavior.
operationally defined as psychosis appearing or not appearing as the final diagnosis in the child’s clinical files.

4.32 Family variables

There were three family variables; two were socio-demographic and one was clinical.

The two socio-demographic family variables were socio-economic status and family structure.

Family socio-economic status was determined by rating each parent’s occupation, which is routinely requested on the CBCL form, on Daniel’s scale of prestige (1983). For families with more than one parent, the more prestigious score was used. The score of 10 denotes highest prestige (eg. professor) with the jobless scored 80.

Determining family structure involved, first of all, examining the range of family types represented in the study sample. They included the following: two natural parents, single mothers, single fathers, one natural parent and one step parent, two adoptive parents, two foster parents, and two grandparents. The categorisation of this data for analysis was guided by an interesting finding from the Mater-University of Queensland Study of Pregnancy (MUSP), a current Australian longitudinal study of over 8000 women recruited during their attendance at a Brisbane pre-birth clinic. It has been reported that the rate of behaviour problems in their children at 5 years of age, on a shortened form of the CBCL, was independently associated with (a) the level of conflict in the relationship between the mother and her partner, and (b) the mother changing partners (Najman, Behrens, Anderson, Bor, O’Callaghan & Williams, 1997). The former is not unexpected but the latter is of special interest. It suggests that child behaviour
problems may be associated with family structure independently of the quality of family functioning – “mothers who experienced no partner changes (married and single) reported the lowest rates of child behaviour problems” (p. 1357).

It is difficult to compare these results with results from other studies as (a) the family structures being studied, (b) the age of the children, and (c) the other family variables controlled, vary in important ways. For example, Nicholson, Fergusson & Horwood (1999) reported their analysis of data collected from the Christchurch Health and Development Study, a cohort of 1265 children born in Christchurch, New Zealand, pertaining to the association between psychosocial outcomes at age 18 and the children living in a stepfamily. It was concluded that “much of this association appeared to be spurious, and arose from confounding social, contextual, and individual factors that were present prior to the formation of the stepfamily” (p. 405). However, the effects of living in a stepfamily was restricted to those entering such a situation for the first time between age 6 and 16, and the “confounding” family variables controlled for in their analyses included the child having lived in single-parent family prior to age 6, the child’s conduct and attentional problems at age 6, and other negative child and family factors which were more prominent in the “stepfamily” group than the comparison group. Statistical adjustment to neutralize the complex chain of disadvantages associated with being a child who enters a stepfamily situation - including factors such as having lived for a time in a single-parent family which might be expected to be intrinsic to the process of becoming a member of a stepfamily - raises questions about the meaningfulness of the adjusted outcomes.

Similar findings are reported by Dunn, Deater-Deckard, Pickering, O’Connor, Golding and the ALSPAC Study Team (1998) on data collected in the Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC), a longitudinal study of over 7000 children born in Avon, England, in
1991 – 1992. They reported that differences between step, single-parent and “non-stepfamily” households with regard to a 4 year old child’s strengths and difficulties (assessed with the Strength and Difficulties Questionnaire) became insignificant after controlling for child and family factors such as the mother’s negative attitude to the child (sample item: “This child gets on my nerves”). Again, it needs to be asked if a meaningful analysis of the effect of being a member of a step-family can take place when negative family factors which might be associated with it are removed. Interestingly, in this study the number of previous live-in partner relationships that the mother had experienced prior to the birth of the target child had been recorded. While there was no effect on the outcomes for the 4 year olds, there was a negative effect on the target children’s older siblings, who averaged 7 - 8 years; the older sibling, unlike the target child, would have had direct exposure to the transitions in their mothers’ live-in relationships. It is suggested by the authors that of two alternative explanations for the association between child problem behaviours and the number of non-father partners to their mother in the home – the direct impact of non-father males entering the child’s home, or the characteristics of mothers who undergo these transitions – the evidence favours the former.

However, alternative interpretations of the different results for the 4 year olds and the 7-8 year need to be considered. The adverse effects of exposure to either of the hypothesised negative factors may not manifest in problem behaviours until children are of school age. This may be due, for example, to (a) the child being unable to meet the developmental demands of transition to school, (b) the mother or male partner being unable to meet the child’s needs at this time, (c) the family structure (for example, the stability in the mother-partner relationship) being unable to meet the child’s needs at this time, and so forth. That is, the negative effects may be cumulative within the child and therefore associated with dose and duration of exposure; the
negative effects may have impact through a “critical” period in the child’s emotional and social development when their specific requirements are not met (Skynner, 1976); or both.

While the mechanisms which underlie the negative association between membership of a stepfamily and the child’s psychosocial outcomes are not yet clear from the available lines of research, the literature does suggest that this factor is associated with disadvantages independent of the other two family variables included in the present study: socio-economic status and family functioning. These considerations influenced the decision to consolidate the seven categories of family structure found in the present study sample into the following three: “two natural parents”, “step-families”, and “other”.

The clinical family variable was family functioning. The Family Assessment Device (Version 3) (Miller, Epstein, Bishop & Keitner, 1985) used for this purpose was one of several instruments for obtaining ratings from individuals on the functioning of their family in clinical and research in use at the time of the commencement of the study. It was selected on the basis of the encouraging early evidence of its psychometric properties, and the perceived clinical relevance of the dimensions of family functioning derived from it. It has since been used in numerous epidemiological and clinical studies both in Australia and internationally, including the Western Australia Child Health Survey (Silburn, Zubrick, Garton, Gurrin, Burton, Dalby, Carlton, Shepherd & Lawrence, 1996).

As can be seen on Table 8, the test-retest reliability at one week was little more than adequate given that the dimensions of the MacMaster model of family functioning the FAD attempts to

---

47 The statistic used is not stated
measure are regarded as stable and enduring characteristics of a family, at least prior to intervention.

<table>
<thead>
<tr>
<th>scale</th>
<th>correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Solving</td>
<td>.66</td>
</tr>
<tr>
<td>Communication</td>
<td>.72</td>
</tr>
<tr>
<td>Roles</td>
<td>.75</td>
</tr>
<tr>
<td>Affective Responsiveness</td>
<td>.76</td>
</tr>
<tr>
<td>Affective Involvement</td>
<td>.67</td>
</tr>
<tr>
<td>Behavior Control</td>
<td>.73</td>
</tr>
<tr>
<td>General Functioning</td>
<td>.71</td>
</tr>
</tbody>
</table>

Table 8. Test-retest reliability on the FAD (Version 3) (from Miller, Epstein, Bishop & Keitner, 1985, p.347)

With regard to the validity of the FAD, there is evidence to suggest that the scores are consistent with ratings of family functioning by experienced family therapists. It is reported that on 6 of the 7 scales, families of psychiatric patients rated by a clinician as unhealthy on a given dimension had mean family FAD scores which were significantly higher (indicating poorer perceived functioning) than families rated as healthy ($p<.05$)\(^{48}\). This was an encouraging finding given the small number of subjects involved ($N = 36$ families). Furthermore, there is some evidence to suggest that answers on the FAD are not strongly influenced by social desirability.

\(^{48}\) The exception was Behavior Control. The difference between the groups, however, was in the expected direction.
However, the attempt to find evidence of convergent validity in the relationship of its scales to those of two other family self-rating scales, the Family Unit Inventory and FACES II, led to somewhat ambiguous results (Miller, Epstein, Bishop & Keitner, 1985). While there are inherent limitations in a methodology where it is assumed, for example, that families with serious problems in “communication” and “affective responsiveness” can and will accurately report their difficulties on a pencil and paper test, what families indicate about themselves is clearly of interest.

While the full 60 items of the FAD (Version 3) were individually administered to parents at registration, the 12 item General Functioning score was used in the present study\(^{49}\). For children with more than one parent their FAD scores were averaged to provide one family functioning score for each case.

4.33 Treatment variables

Finally, there were three treatment variables; length of stay, the program to which the child was admitted (ie. Junior, Middle or Senior program), and whether the admission took place during the earlier or later of the two distinct “institutional eras” of Arndell Child and Adolescent Unit during the 1990s.

Length of stay was calculated as the period between admission of the child to either residential or day treatment and their discharge from the Unit.

\(^{49}\)There is no convenient “total score” for this instrument as it measures six differing constructs.
As for the program into which the child was admitted, this reflects more than the age by which the child is allocated. During the period of the present study, the Middle school was the only program with a residential component, representing continuity with the residential philosophy of the Unit at its inception. It was decided that it was impractical to sub-categorise children in the Middle school into day and residential treatment groups for the present study as there was no absolute distinction between these categories; most children in the Middle school would spend a period as a resident and also as a day patient during their period of admission. There were multiple determinants of whether a child was treated as a resident or a day patient, including the pragmatics of travel (it was obviously impractical for children from regional NSW to commute on a daily basis, for example), individual clinical needs (for example, residential children approaching discharge were “weaned” into day status, then often integrated into their new schools several days a week) and the needs of the Unit (a particularly demanding child would not be brought into residence if, for example, the staffing level was felt to be overstretched, or the mix of children at the time felt to be too volatile). Most children in the Middle program felt that participation as a resident was desirable – not surprising given the variety of sporting and leisure activities and outings on offer “after school” - and spent at least some time in residence, even if it was only several weeks during the “holiday programs” that would take place between school terms.

The earlier “institutional era” during the 1990s was marked by its adherence to the then “traditional” principles of psychodynamic milieu therapy. Psychodynamic milieu therapy was the dominant therapeutic trend in intensive day and residential treatment units for severely emotionally disturbed children in the Anglosphere from the 1950s to the 1970s (Aichorn, 1935; Redl & Wineman, 1951, 1952; Trieschman, Whittaker & Brendtro, 1969). The later “institutional era” was marked by a distinct and explicit departure from and rejection of these
principles, and the adoption of “empirically based” treatments with distinctly behavioural and
cognitive-behavioural principles which became dominant in the 1980s (Kazdin, Esveldt-

As mentioned [4.2], an important feature of the Unit from its inception in 1958 and a feature still
evident at the commencement of the evaluative research program in 1990, was the full
integration between mental health and education staff at the Unit. This was formed around a
strong personal and professional alliance between the director of Arndell Unit and the principal
of Arndell SSP (School for Special Purposes), and a shared investment in the value of the
therapeutic milieu as the primary agent of treatment. In the second part of the decade, there was
a change in the relationship between the mental health and the education components, with a
deliberate move away from the milieu model by new appointees for the key positions of unit
director, school principal and nursing manager. This lead to an operational separation between
the two components, leading to separate meetings for the staff, and a downgrading of whole
Unit staff meetings. This was accompanied by a change in the education staff’s focus on the
child’s problem behaviours in the classroom seen in the context of intimate knowledge of
difficulties in their emotional development and family issues, to a more conventional focus on
academic goals and behaviours appropriate for regular classroom settings.

As for the clinical staff, there was a more direct, ahistorical focus on behavioural issues in the
work with the individual child and with their family in the second part of the decade, and there
was less interest in the social interactions of the child in the milieu as an important source of
understanding their emotional needs and difficulties. Moreover, there was no longer interest in
examining conflicts and tensions in the milieu - including staff-client relations and staff-staff
relations, as well as the general mood of the Unit - as reflecting important issues within the children.

This represented a specific turning away from the psychodynamic-systems approach of Redl (Redl & Wineman, 1951, 1952) in the U.S. and the Bion-Tavistock developments in the U.K. (Bion, 1961; Trist & Murray, 1990) which informed the milieu therapy since the early days of the Children’s Unit. Along with the reorientation towards an explicitly behavioural approach in the classroom, the psychodynamically oriented therapists strongly identified with working with the dynamics of the milieu came to be replaced with cognitive-behaviourally oriented therapists trained in working with individuals and families but with little interest in small or large group dynamics.

This represented a dramatic shift in the dominant figure-ground configuration in the Unit. What was previously “signal” – for example, problem behaviours in the classroom which was perceived as reflecting distress or conflicts in the child, distress or conflicts in the relationship between members of the Unit milieu, or distress or conflicts in the child’s family - was now considered “noise”. That is, problem behaviours in the classroom were seen as a way for the child to distract education staff attention from his difficulties in sitting at a desk and facing his school work. In sharp contrast to the milieu therapy orientation of the first part of the 90s, the involvement of education staff in (what came to be seen as seemingly endless) discussion with the child, class members, and clinical staff attempting to work out the “real”, “internal” causes of the child’s behaviour at that time, was perceived as reinforcing the child’s off-task behaviour. The emphasis came to be placed on returning the child to on-task behaviour while leaving any other issues to be explored outside of class room time in designated group therapy, individual therapy, or family therapy times.
It would be misleading, however, to suggest that there was great homogeneity among staff at any time with regard to therapeutic orientation, or their core beliefs and attitudes with regard to the active ingredients of the treatment (and educational) efforts in which they were engaged. For example, issues such as the degree to which teaching staff were engaged in core educational instruction as opposed to attempting to meet the emotional needs of the child, or the boundaries between the roles of clinicians, nurses and education staff, remained lively and controversial topics throughout the entire period of the present study. Yet, in delineating the difference between the earlier period of the evaluative research program during which psychodynamic milieu therapy was practiced, and the later period when deliberate changes in policy, senior staffing, and Unit structure were made towards the “empirically-based” therapies of its times, it would not be an exaggeration to state that a quite deliberate, abrupt, revolutionary change occurred at Arndell Unit in the mid 90s.

The operational definition of the time spent in the “old” versus “new” institutional era for each child was the proportion of their admission spent before 1 June 1996, which ranged from 0 (admission entirely in the old era) to 1 (admission entirely in the new era), with intermediate values for children whose admission crossed over 1 June 1996\(^50\).

4.34 Outcome variable

Outcomes were measured by the Total Problem Score on the Child Behavior Checklist completed by parents 6 months and 12 months after discharge.
While efforts were made to obtain outcome data from parents as close as possible to 6 months and 12 months after discharge, there was some variation in the promptness with which the CBCLs were completed and returned. In some instances, the parents fell behind in their completion of forms during the admission period\textsuperscript{51} and these may not have been returned until well after the child’s discharge. Therefore, the following strategy was adopted. Using the date of completion of the CBCL as indicated by the raters on their form, the time which had elapsed between discharge and completion of protocol was calculated. Those completed within the range of 3 months to 9 months post-discharge were classified as “6 months outcomes”, and those completed within the range of 9 months to 15 months post-discharge were classified as “12 month” outcomes. For cases where there were two protocols available from the same rater falling into the same range, the protocol completed closer in time to the 6 month and 12 month post-discharge dates was selected.

In earlier discussions of the CBCL in Chapter 2, it was argued that the data obtained from its administration represented the product of a complex procedure that involves a \textit{communication} from the rater regarding their \textit{perception} of the subject's behaviour, which was vulnerable to distortions, biases and demand characteristics operating on the rater at either point.

As for \textit{communication}, the following steps were taken to insulate the administration of the CBCL from the demands of the clinical process in which the parents were intensively involved. A member of the Unit’s clerical staff managed the collection of data from families. This individual had no previous background in research but possessed the administrative and

\textsuperscript{50} This date represents the commencement of E as unit director (Table 29), who was appointed with a clear intention to enact change.

\textsuperscript{51} The Arndell Evaluative Research Program involved routine data collection during the course of treatment. The present study involves a subset of the accumulated data selected to address the specific research question.
“people” skills often found in the front offices of mental health clinics. She was well motivated to take up part-time duties in what evolved into the specific role of evaluative research program administrator, in addition to her continuing part-time role as one of two receptionists at the Unit. The administrator distributed and collected the forms, in person, from families at admission and during admission at the Unit, and followed up by phone and post after they left the Unit. In this way the clinical staff, with whom the parents usually developed very intense relations, were bypassed. On the occasions when the administrator detected that slowness in returning a form was associated with a parent’s negative experience in the course of treatment of the Unit, she was able to sympathise and to encourage them to “return the form anyway” as it was “important for us to know about all outcomes, both good and bad”. The relationships she developed with often very hostile or distressed parents as they attended and left the Unit – sometimes complaining to her about the terrible therapists they had to see! - helped to sustain a friendly down-to-earth working alliance despite the vicissitudes of the treatment process.

As for whether the ratings reflected an “accurate” perception of the child, the focus of the Unit on working directly with family dynamics meant that whether a change in ratings was due to (a) changes in the child’s behaviour, (b) changes in the parent–child relationship, or (c) changes in the perceptions of the parent, was not relevant. It was appropriate in a multi-modal child and family treatment unit to accept any decrease in parental reports of total child problem behaviour as achievement of a significant goal.

The adoption of the Total Problem Score rather than the syndrome subscales as the outcome variable reflects the problem of assuming the clinical significance of the direction of change in subscales, as discussed [2.42]. To continue with the earlier example, while the overcontrolled youth who shows therapeutically desirable increase in assertiveness may show an increase on
the Aggressive Behaviour syndrome score, genuine therapeutic change is likely to lead also to a
decrease in internalizing behaviours (rated on the Withdrawn, Somatic Complaints and the
Anxious/Depressed syndrome scores) and therefore to an overall decrease in the Total Problem
Score. At the same time, as a composite of specific behavioural observations, the Total Problem
Score offers a measure likely to be less vulnerable to biases arising from rater expectations in a
naturalistic setting than global or general judgments about behaviours (Kent & Foster, 1977) or
judgements as to whether a child is “better” at a particular time period when compared with
recollections of an earlier period (Lambert, Salzer & Bickman, 1998).

Although it was routinely collected, data from the Teacher Report Form were ruled out as an
outcome variable for the present study due to its poor reliability when the same rater is not
available.52

Self-reports by the children and adolescents on the Youth Self Report (Achenbach’s parallel self
report form), also routinely collected, was also not considered as an outcome variable for the
present study due to (a) the instrument only being suitable for a portion of the children and
adolescents admitted to the Unit (i.e. youths 11 and older), (b) the poor compliance of the
children and adolescents, and (c) the strong interaction between the treatment progress and the
youth’s willingness to communicate their difficulties. That is, it was our experience that angry,
alienated youths distrustful of adults and institutions would not reveal many of their difficulties
and vulnerabilities on such an instrument at admission (often scoring 0 on Total Problem Score),
but became more prepared to reveal these aspects of themselves as they developed therapeutic
rapport; it was therefore expected that for a significant number of youths a good therapeutic
outcome would be associated with an increase in Total Problem Score on the YSR. While this
form of communication can be readily understood by the clinicians involved with the youth, in the absence of validity scales such as those in the MMPI-2 it was not possible to devise a simple operational procedure to sort out “genuine” from “non-genuine” YSR protocols.

It was anticipated that there would be clinically meaningful differences between parents in their ratings, and that these might be systematically associated with family structure and the clinical features of the presentation. In line with the present focus on outcomes rather than parental differences, the following procedure was put in place. Rather than leave to each family to work out which parent might complete the checklist, systematic differences were controlled through administering the CBCL to each parent separately (where there was more than one parent), and pointing out to them the standard instructions which appear on the form to “fill out this form to reflect your view of the child’s behaviour even if other people might not agree”. While the terms “mother” and “father” will be used, these terms represent a parent or parent-surrogate of female and male gender, respectively. For children with two “parents” their CBCL scores were averaged to provide one set of parental CBCL scores for each case.

4.35 The schedule for the administration of instruments

The variables and instruments in the present study and the schedule for their administration are summarised in Table 9.

52 Given the long duration of treatment at the Unit, it was rare for children to return to the same teacher following discharge.

53 The original US normative study, for example, involved the test being administered to any parent or parent surrogate who agreed to participate in response to a door knock (Achenbach, Hensley, Phares & Grayson, 1990) [2.51].
<table>
<thead>
<tr>
<th>predictors</th>
<th>outcomes</th>
<th>variable</th>
<th>instrument</th>
<th>rater/source</th>
<th>T0</th>
<th>T1</th>
<th>T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>child socio-demographic</td>
<td></td>
<td>gender</td>
<td>clinical file</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>age</td>
<td>clinical file</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>child clinical.</td>
<td></td>
<td>home behaviour</td>
<td>CBCL</td>
<td>parent</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>school behaviour</td>
<td>TRF</td>
<td>teacher</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>family socio-demographic</td>
<td></td>
<td>socio-economic status</td>
<td>Daniel’s Scale</td>
<td>receptionist</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>family structure</td>
<td>FAD</td>
<td>parent</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>family clinical</td>
<td></td>
<td>family functioning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatment</td>
<td></td>
<td>length of stay</td>
<td>clinical file</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>program</td>
<td>clinical file</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>pre versus post 1 June 1996</td>
<td>clinical file</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(T0 = admission, T1 = 6 months post-discharge, T2 = 12 months post-discharge)

(X indicates when ratings take place)

**Table 9.** The schedule for the administration of rating instruments
4.4 The criteria for inclusion in this study

The selection of subjects for the present study followed these criteria:

1. The systematic data collection, which commenced in August 1990 at the Unit was terminated on 25 September 2001. This meant that there were children at the termination date who had been enrolled in the Arndell Evaluative Research Program (i.e. from whom data had been collected at pre-treatment or during treatment) but for whom data collection at either or both post-discharge follow-up points was yet to take place. It was able to be determined that all children admitted to inpatient or day treatment up to August 1999 had been followed up at both T1 and T2. That is, it was only after this enrolment date that there were families of children enrolled in the study not yet contacted for T2 data collection or who had been contacted but yet to return the follow-up CBCL. It was considered desirable to avoid systematic bias associated with parental delays in returning the forms (which may be associated with parental dissatisfaction with treatment or with the outcome achieved for their child), or systematic bias associated with length of stay (i.e. parents of children with longer lengths of stay were contacted for post-discharge followup at a later date than parents of children with shorter lengths of stay). This was ensured by only including children for whom pre-treatment data was collected before the end of August 1999.

2. Each subject was required to have at least one parent who completed the CBCL at registration and also at either the 6 month or the 12 month post-discharge follow-up.

3. Each subject was required to have a TRF completed by their previous teacher (i.e. their teacher immediately previous to their admission to Arndell).
This resulted in 159 subjects. There were 133 subjects for whom 6 month outcomes were available, 134 subjects for whom 12 month outcomes were available, and 108 subjects for whom both 6 and 12 month outcomes were available.

Some data is available on participation rates in the study relative to the population of children attending treatment over this period of time. A review was made of the 102 consecutive admissions to the Unit in the three-year period between August 1993 and the end of July 1996. It was found that 60 children met the minimum criteria for inclusion, while 42 did not. An analysis of the 42 for whom we did not have a minimal data set is presented in Table 10.

<table>
<thead>
<tr>
<th>5</th>
<th>Failed to complete CBCL at registration</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Completed CBCL at registration, but declined to participate further in the study.</td>
</tr>
<tr>
<td>10</td>
<td>Completed CBCL at registration and during admission, but declined to participate in the post-discharge follow-up</td>
</tr>
<tr>
<td>7</td>
<td>Completed CBCL at registration and during admission, and agreed to participate in the follow-up, but failed to return forms</td>
</tr>
<tr>
<td>10</td>
<td>Miscellaneous (eg. forms completed by different individuals at registration and at follow-up due to change in child’s living arrangements)</td>
</tr>
</tbody>
</table>

**Table 10.** An analysis of the 42 children admitted August 1993 to end July 1996 where a “minimum” data set was not obtained.

As seen in Table 10, we failed to obtain protocols at the outset from the parents of 5 children. This probably reflects in some part failure to engage the cooperation of the family’s case-manager, as families who were – due to distress or time pressures – unable to complete the
forms at registration were followed up by the research administrator. After this point, parents were presented with informed consent forms with regard to their participation in the study. As mentioned [4.21], the form was designed in such a way that they could agree to participate in the study, yet disagree to have their child’s school approached. As seen in Table 10, the parents of 10 children declined to participate at all. If they agreed to participate, they were presented with another informed consent form at discharge to seek consent to follow-up. As seen in Table 10, the parents of a further 10 children dropped out at this point. The parents of 7 children agreed to participate yet failed to return forms.

The miscellaneous group of ten includes the not insignificant number of children who at follow-up were living separately from the adult or adults who had completed the forms at registration.

We note here that, as mentioned, it was made clear to parents at the outset that the clinical services to be offered them would not be affected in anyway by their decision to participate or not participate in the study. While consistent with ethical considerations, it could be argued that a less liberal approach may have increased the level of participation among the parents. However, the advantage of our approach was that the data was less likely to be compromised by poor quality ratings produced by poorly motivated participants.

It is likely, then, that those meeting the criteria for inclusion in the present study are not representative of all those admitted to the Unit during this time period. Both selection and mortality (“drop out”) factors are likely to be relevant, and to interact with outcomes. Those meeting the criteria for inclusion in the present study are likely to be children with more cooperative and/or compliant parents, which is in turn likely to be influenced by their child’s outcomes. Similarly, those returning questionnaires at 12 months post–discharge may differ
from those not returning questionnaires at this time with regard to the outcomes for their child, reflected in outcomes at 6 months post-discharge.

At the same time, it needs to be pointed out that the representativeness of the study sample does not have quite the same relevance to the present research question as it does to research aiming to address treatment effectiveness. While, as recognised, the selection and mortality factors may be expected to inflate the outcome for the study sample over the population of all children attending treatment, as discussed [3.61], there is no reason to expect that the relationship between predictor variables and outcome is different for the study group and the broader treatment population.
4.5 Research design and aims

The research question called for a multiple linear regression analysis of the relationship between the selected predictors and the outcome variable. This well-developed procedure, familiar in epidemiological and other naturalistic studies, is able to assess the relationship between each predictor variable and the outcome variable, while holding constant the influence of the other predictor variables on the outcome variable.

The present design, however, aimed to avoid the common types of methodological weaknesses in analyses of this type, as outlined in Chapter 2.

First, the use of measures of morbidity both as a predictor and as an outcome leads to the clinically uninformative finding that the strongest predictor of post-treatment morbidity is pre-treatment morbidity, as has already been discussed [2.51].

Second, as discussed [2.52], while the typical use of behaviour rating scales assumes that differences between raters, providing ratings of behaviour over the same time period, reflect measurement error, differences are often likely to reflect clinically significant, systematic variation. The dimension of pervasiveness of a child’s behaviour, represented in problem behaviours being reported to take place both at home and at school, is logically distinct from the dimension of the severity of problem behaviours in either setting. It is suggested that the clinical significance of these two different dimensions of problem behaviours is quite distinct.

Third, while the application of the same standardised general behaviour rating scale (the CBCL) to all subjects in a study has the advantages of uniformity, and the advantage of a focus on Total Problem Score over the narrower syndrome subscales has already been discussed, there is a
limitation to the use of the Total Problem Score with particular types of children. Children with circumscribed difficulties, even if they are very severe, will show a low total problem score on the CBCL. For example, a child presenting with elective mutism which has been resistant to treatment for several years, but with no other *behavioural* disturbance observable to the parents, may well score in the “normal” range. Furthermore, starting from such a low base-line a good therapeutic result would be barely detectable (i.e. the floor effect)\(^{54}\). For these reasons, the pooling together of CBCL data from children with broad ranging difficulties and those with circumscribed symptoms is inappropriate and likely to lead to obscuring of the treatment effect.

Fourth, more generally, an attempt to fit one multiple regression model of predictors for the entire study sample is likely to lead to obscuring of possibly quite different interactions between variables taking place in different types of children. Even leaving aside the issue of floor effects, it does not seem reasonable that the variables, and interaction between variables, leading to a good outcome in a child with a broad range of extreme problems behaviours both at school and at home, would be the same for a child with very circumscribed problem behaviours. Clinically meaningful results about predictors of outcome are more likely to emerge if the analysis proceeds in a way which allows clinically distinct groups to show varying results. This is consistent with the conclusion from the literature review pertaining to the importance of conducting analyses within clinically meaningful subgroups [3.63],

These four considerations suggest a departure from conventional practices with regard to the use of the CBCL data in multiple regression analyses of predictors of treatment outcome is required.

\(^{54}\) The standardised implementation of a CBCL-based evaluative research program is often resisted by clinicians in many settings due to this necessarily “blunt” characteristic of the instrument. They have good cause to fear that their clients would appear not very disturbed and that they would be exposed to the criticism that their efforts are being misdirected to the “worried well”, or that their work would be deemed ineffective.
These considerations lead to a data analytic strategy where a multiple regression analysis takes place within four clinically meaningful subgroups constructed according to pre-treatment ratings by parents and teachers, as follows (Table 11).

<table>
<thead>
<tr>
<th>Moderate parent-reported problems</th>
<th>Moderate teacher-reported problems</th>
<th>Severe parent-reported problems</th>
<th>Severe home problem subgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODERATE PROBLEM subgroup</td>
<td>SEVERE SCHOOL PROBLEM subgroup</td>
<td>SEVERE-PERVERSIVE PROBLEM subgroup</td>
<td></td>
</tr>
</tbody>
</table>

Table 11. The formation of the four subgroups

The first subgroup consists of children receiving relatively moderate total problem scores from both parents and teachers. This will be referred to as the moderate problem subgroup subgroup. The second subgroup consists of those children reported by teachers to show relatively severe problem behaviours but reported by parents to show relatively moderate problem behaviours - the severe school problem subgroup. The third group would consist of those reported by parents to showing relatively severe problem behaviours but reported by teachers to show relatively moderate problem behaviours - the severe home problem subgroup. The fourth group would then consist of those children reported by both parents and teachers to show relatively severe problem behaviours - the severe-pervasive problem subgroup.
The construction of these clinically meaningful subgroupings within which multiple linear regression analyses are performed addresses the common methodological weaknesses of studies of this type, as identified. The differences between parent and teacher ratings are treated as conceptually meaningful, and form the basis for the construction of the four subgroups. Pre-treatment pathology is not included among the predictors; rather than overwhelming the correlation between other predictors and outcome, the differences in pre-treatment pathology is accounted for in the constant for each subgroup. The subgrouping offers a way to avoid inappropriately aggregating the Total Problem scores of children with circumscribed problems with the Total Problem scores of children with pervasive problems. By allowing the analysis to proceed in a way which allows the clinically distinct subgroups to show varying results, it maximises the possibility of clinically meaningful results about predictors of outcome to emerge.

At the same time the subgrouping does not lead to reduction of power; this strategy retains the statistical advantage of utilising data from the larger group of subjects for the purposes of calculating estimates and errors, and in filling missing values.56

55 Naming this the “low problem” subgroup was not appropriate as no child admitted to the Unit could be considered to have a “low” level of problems.
56 It might be asked why the present procedure was preferred over other commonly used statistical procedures. Procedures such as logistic regression and discriminant analysis following the conversion of the outcome data into categories of “treatment successes” and “treatment failures” were rejected because of the loss of information contained in the CBCL total problem scores (this is discussed further in 5.4). Structural equation modelling was deemed inappropriate because there were no factor-analytic measurement aspects to the problem. It might also be asked why change scores, rather than endpoint scores, were not used as the outcome measure. As outlined, the use of pre-treatment parent ratings was integral to the formation of the four subgroups. Change scores were not appropriate as an outcome measure due to the artificial association which would exist between predictor and outcome in this case. Any error of measurement in the pre-treatment variable would be reiterated negatively in the dependant variable. It would be possible, for example, to find a strong but entirely spurious relationship between the predictor and the change score (Cronbach & Furby, 1970).
It is not argued that this data strategy represents a new convention which might generally be preferred over any other convention. Nor is it argued that there are no other questions which might be asked of this data set. Moreover it is not implied that the present study in any way exhausts the clinical and scientific value to be mined from the wealth of data which quickly accumulates once an ongoing evaluative research program is initiated in a busy clinical setting.

Furthermore the role of luck in the present study needs to be acknowledged. It has been argued [2.21] that it was important to distinguish between two different types of questions which commonly arise in treatment evaluation – those pertaining to treatment effectiveness and those pertaining to predictors of treatment outcome – as the design requirements were quite different. The question addressed in the in the present study pertains to predictors of treatment outcome. As mentioned [3.62], as a naturalistic study, it was entirely fortuitous that the dominant treatment type changed abruptly during the decade that evaluative research was taking place at Arndell Child and Adolescent Unit, so that treatment type was able to be included among the predictors of outcome.

### 4.51 Missing data

The criteria for selection into the present study, as outlined, minimised the problem of missing data.

There was only one predictor variable with missing observations. Of the total of 159 subjects, the family functioning data was unavailable for 11 subjects. The strategy was adopted of filling in the missing values with estimates derived from performing multiple linear regression on the other predictor variables. While the proportion of variance on the family functioning score
accounted for by this method was small (adjusted $R^2 = .01$), the method was at least free of bias with regard to the main research question (ie. predictors of outcome) and allowed the main statistical analyses of interest to proceed with the full number of subjects.

As stated, 6 month outcomes were available for 133 cases and 12-month outcomes were available for 134 cases, and there were 108 cases with both 6 month and 12 month outcomes. (By the nature of the selection criteria, there were no cases in this study missing both 6 month and 12 month outcomes). As there was interest in differences, if any, between predictors of 6 month outcomes and predictors of 12 month outcomes, no steps were taken to fill missing 6 month values from 12 month values, or vice versa, so that it remained possible to perform separate analyses for these two time periods without contamination.
4.6 Summary of study variables

The following table (Table 12) provides a summary of the classificatory, predictor and outcome variables of this study.

<table>
<thead>
<tr>
<th>CLASSIFICATION VARIABLES</th>
<th>INSTRUMENT</th>
<th>ADMINISTRATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>parental rating of problem behaviours</td>
<td>Total Problem Score on CBCL</td>
<td>admission</td>
</tr>
<tr>
<td>teacher rating of problem behaviours</td>
<td>Total Problem Score on TRF</td>
<td>admission</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PREDICTOR VARIABLES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>child socio-demographic</td>
<td>clinical file</td>
</tr>
<tr>
<td>gender (F = 0, M = 1)</td>
<td>admission</td>
</tr>
<tr>
<td>age (years)</td>
<td>admission</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>child clinical</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>externalizing/internalizing index</td>
<td>index derived from CBCL</td>
</tr>
<tr>
<td>(range = 0 – 1, higher = predom. extern)</td>
<td>admission</td>
</tr>
<tr>
<td>diagnosis of psychosis</td>
<td>clinical file</td>
</tr>
<tr>
<td>(not psychotic= 0, psychotic = 1)</td>
<td>discharge</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>family socio-demographic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>socio-economic status</td>
<td>Daniel’s Scale of</td>
</tr>
<tr>
<td>(range 10 – 80, lower = more prestige)</td>
<td>Occupational Prestige</td>
</tr>
<tr>
<td>family structure: two natural parents</td>
<td>clinical file</td>
</tr>
<tr>
<td>(no = 0, yes = 1)</td>
<td>admission</td>
</tr>
<tr>
<td>family structure: step family</td>
<td>clinical file</td>
</tr>
<tr>
<td>(no = 0, yes =1)</td>
<td>admission</td>
</tr>
<tr>
<td>family structure: other</td>
<td>clinical file</td>
</tr>
<tr>
<td>(no = 0, yes =1)</td>
<td>admission</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>family clinical</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>family functioning</td>
<td>General Functioning Scale of</td>
</tr>
<tr>
<td>(lower = better functioning)</td>
<td>the Family Assessment Device</td>
</tr>
<tr>
<td>treatment</td>
<td></td>
</tr>
<tr>
<td>length of admission (months)</td>
<td>clinical file</td>
</tr>
<tr>
<td>program: Junior School (no = 0, yes =1)</td>
<td>discharge</td>
</tr>
<tr>
<td>program: Middle School (no = 0, yes =1)</td>
<td>discharge</td>
</tr>
<tr>
<td>program: Senior School (no = 0, yes =1)</td>
<td>discharge</td>
</tr>
<tr>
<td>proportion of admission after 1 June</td>
<td>clinical file</td>
</tr>
<tr>
<td>(none = 0, all = 1)</td>
<td>discharge</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OUTCOME VARIABLES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>parental rating of problem behaviours</td>
<td>Total Problem Score on CBCL</td>
</tr>
<tr>
<td>parental rating of problem behaviours</td>
<td>6 months post-discharge</td>
</tr>
<tr>
<td>parental rating of problem behaviours</td>
<td>Total Problem Score on CBCL</td>
</tr>
<tr>
<td>parental rating of problem behaviours</td>
<td>12 months post-discharge</td>
</tr>
</tbody>
</table>

Table 12. Summary of study variables
CHAPTER 5.
RESULTS OF STATISTICAL ANALYSES OF DATA

5.1 Introduction to this Chapter

This chapter describes the results of statistical analyses of the data according to the methods and aims outlined in the previous chapter.

These preliminary steps are outlined: the four subgroups are constructed [5.2]; the predictor variables are examined for the study sample as a whole and for the subgroups [5.3]; the outcome at 6 months and 12 months post-discharge are examined for the study sample as a whole and for the four subgroups [5.4].

Finally, the selected predictors are tested within each of the four subgroups for their relationship to outcome at 6 months and 12 months post-discharge [5.5].
5.2 The formation of the clinical subgroups

Four groups of similar size were formed by setting the cutoff between “moderate” and “severe” parent–reported problems (CBCL Total Problem Score) at 74, and the cutoff between the “moderate” and “severe” teacher-reported problems (TRF Total Problem Score) at 83, as seen in Table 13.

<table>
<thead>
<tr>
<th></th>
<th>moderate parent-reported problems</th>
<th>severe parent-reported problems</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(CBCL Total Problem Score &lt; 74)</td>
<td>(CBCL Total Problem Score ≥ 74)</td>
</tr>
<tr>
<td>moderate teacher-reported problems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(TRF Total Problem Score &lt; 83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MODERATE PROBLEM subgroup (n=42)</strong></td>
<td>mean CBCL Total Prob Score = 52.12 (s.d 13.6)</td>
<td>mean CBCL Total Prob Score = 90.35 (s.d 13.78)</td>
</tr>
<tr>
<td></td>
<td>mean TRF Total Prob Score = 50.76 (s.d 22.23)</td>
<td>mean TRF Total Prob Score = 51.24 (s.d 24.94)</td>
</tr>
<tr>
<td>severe teacher-reported problems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(TRF Total Problem Score ≥ 83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SEVERE SCHOOL PROBLEM subgroup (n=37)</strong></td>
<td>mean CBCL Total Prob Score = 55.97 (s.d 13.8)</td>
<td>mean CBCL Total Prob Score = 102.57 (s.d 21.61)</td>
</tr>
<tr>
<td></td>
<td>mean TRF Total Prob Score = 109.35 (s.d 20.90)</td>
<td>mean TRF Total Prob Score = 112.45 (s.d 19.43)</td>
</tr>
</tbody>
</table>

**Table 13.** The formation of the four clinical subgroups.

The box plots of means, standard deviations and range of parent and teacher ratings for the four subgroups are represented in Figure 2.
The cutoffs resulted in very similar teacher ratings for the moderate problem and the severe home problem subgroups, on one hand, and for the severe school problem and the severe-pervasive problem subgroups, on the other. As for the parent ratings, the moderate problem and the severe school problem subgroups are similar, on one hand, and the severe home problem and the severe-pervasive problem subgroups are similar, on the other. The severe-pervasive problem subgroup is shown to consist of cases rated by both parents and teachers as showing severe problems in home and school settings, respectively, with the parent ratings comparable to those for the severe home problem subgroup, and the teacher rating comparable to those for the severe school problem subgroup.

It is concluded that distinct subgroupings consistent with the stated aims were able to be achieved.
5.3 Predictor variables across the clinical subgroups

5.31 Child socio-demographic variables

There are two child socio-demographic variables among the predictors of outcome: gender and age at admission.

The distribution of gender in the study group as a whole and across the four subgroups, and the results of a statistical test of differences in distribution across the four subgroups, is represented in the following table (Table 14a).

<table>
<thead>
<tr>
<th>gender</th>
<th>moderate problem subgroup n=42</th>
<th>severe school problem subgroup n=37</th>
<th>severe home problem subgroup n=37</th>
<th>severe-pervasive problem subgroup n=43</th>
<th>total %</th>
<th>P values</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>31</td>
<td>37</td>
<td>31</td>
<td>35</td>
<td>84%</td>
<td>.014</td>
</tr>
<tr>
<td>female</td>
<td>11</td>
<td>0</td>
<td>6</td>
<td>8</td>
<td>16%</td>
<td></td>
</tr>
</tbody>
</table>

**Table 14a.** Differences in distribution of gender across the four subgroups (chi square)

The overall predominance of males over females (roughly 5:1) is consistent with that reported by other child/adolescent day and inpatient units (as discussed in Chapter 3). However, the predominance of males in these settings is inconsistent with estimates of gender differences in the prevalence of emotional and behavioural problems in the community. Gender differences in the prevalence of psychiatric disorder among children and adolescents appear to depend on the types of problems included under this term (for example, boys are more frequently diagnosed with conduct disorder than girls), their age, the country in which the survey takes place, and other factors about which there is not yet systematic understanding. In a survey of 14 to 21 year olds conducted in Hayfield, Victoria, Australia, the male to female ratio of prevalence of psychiatric disorder was 1: 1.15 (Krupinski, Baikie, Stoller, Graves, O'Day & Polke, 1967). In a
Canadian survey (Offord, Boyle, Szatmari, Rae-Grant, Links, Cadman, Byles, Crawford, Blum, Byrne, Thomas & Woodward, 1987), the male to female ratio among 12 to 16 year olds was the same. In a New Zealand survey of 11 year olds, the male to female ratio was 1.7: 1, but by the age of 15 the same sample showed a greater prevalence of DSM-III disorders among the girls, with a male to female ratio of 1:1.4 (Anderson, Williams, McGee & Silva, 1987). However, McDermott (1996), for example, raises the issue of whether there are gender specific overdiagnoses in particular categories, such as boys being overdiagnosed for conduct disorder. With prevalence rates of up to 41.3% (!) for at least one DSM-III diagnosis among 14 to 16 year olds in one Michigan, USA, survey (Kashani, Beck, Hoeper, Fallahi, Corcoran, McAllister, Rosenberg & Reid, 1987), it is difficult to know how meaningful these gender prevalence estimates are. In the comparison of CBCL scores from 12 cultures (Crijnen, Achenbach & Verhulst, 1997), discussed previously [2.52], gender had a very small effect size (less than 1%) on Total Problem Score for children (6 to 11) with the difference in the direction of higher scores for boys, and no statistically significant effect among adolescents (12 to 17).

The striking inconsistency in male to female ratios between admission rates to child/adolescent day and inpatient units and estimates of community prevalence remains to be explained. It might be suggested that it reflects the reluctance of referring agents, parents, and staff to introduce girls, especially those with internalizing problems, into an environment where they will be exposed to the extreme externalizing behaviours of males. For example, in a recent survey of the 29 inpatient child psychiatric units in the UK (Green & Jacobs, 1998b), the median percentage of diagnosis on admission was 25% for “conduct disorder” and 25% for “mixed disorder of conduct and emotion” – any unit with half the children showing conduct problems is an intimidating place for staff, let alone anxious, depressed or withdrawn girls\textsuperscript{57}.

\textsuperscript{57} This survey appears not to report the male to female ratios, however.
As for differences in distribution of gender across the four subgroups, the $p$ value resulting from chi square is statistically significant. There is an absence of girls in the **severe school problem** subgroup which is difficult to explain. The possibility that this reflects an underlying epidemiological gender difference - that girls do not show significant problems at school without also showing significant problems at home - is not suggested by the available literature. A more likely possibility is that this reflects an aspect of the process which results in the underrepresentation of girls in these Units. The possible explanation of the absence of girls in the **severe school problem** subgroup as simply reflecting the reluctance to bring girls into this environment unless severe levels of home problems “justify” it, is not supported by the fact that females are well represented in the **moderate problem** subgroup. Recognising also that the number of girls in each subgroup is quite small, it is not possible to be more specific at the present time.

The mean and standard deviation of **age** for the four subgroups, and the results of a statistical test of difference in age across the subgroups (one way ANOVA), are presented in the following table (Table 14b).

<table>
<thead>
<tr>
<th></th>
<th>moderate problem subgroup</th>
<th>severe school prob. subgroup</th>
<th>severe home prob. subgroup</th>
<th>severe - pervasive subgroup</th>
<th>P values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>n</strong></td>
<td>42</td>
<td>37</td>
<td>37</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td><strong>mean (s.d)</strong></td>
<td>9.60 (3.38)</td>
<td>9.89 (3.04)</td>
<td>9.97 (2.47)</td>
<td>10.21 (2.32)</td>
<td>.80</td>
</tr>
</tbody>
</table>

**Table 14b.** Differences in age across the four subgroups (one way ANOVA)

These results show no statistically significant differences in age across the subgroups.
There are two child clinical variables among the predictors of outcome: the externalizing/internalizing index and diagnosis of psychosis.

The distribution of the externalizing/internalizing index in the study sample is shown in Fig 3 (below). The values ranged from 0.03 to 0.82, with a lower index representing a predominantly internalizing presentation and a higher index representing a predominantly externalizing presentation. The mean value of the index is 0.54 with a standard deviation of 0.14.

Figure 3 The distribution of the internalizing/externalising index in the study sample.
The mean and standard deviation of the externalizing/internalizing index for the four subgroups, and the results of a statistical test of difference across the subgroups (one way ANOVA), are presented in the following table (Table 14c).

<table>
<thead>
<tr>
<th></th>
<th>moderate problem subgroup n=42 mean (s.d)</th>
<th>severe school prob. subgroup n=37 mean (s.d)</th>
<th>severe home prob. subgroup n=37 mean (s.d)</th>
<th>severe - pervasive subgroup n=43 mean (s.d)</th>
<th>P values</th>
</tr>
</thead>
<tbody>
<tr>
<td>externalizing/</td>
<td>.53 (.14)</td>
<td>.55 (.11)</td>
<td>.53 (.15)</td>
<td>.52 (.14)</td>
<td>.81</td>
</tr>
<tr>
<td>internalizing index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 14c. Differences in the externalizing/internalizing index across the four subgroups (one way ANOVA)

These results show no statistically significant differences across the subgroups. The very close means across all four groups on the externalizing/internalizing index is encouraging as this supports the expectation that this index captures this dimension of child presentation style unconfounded with child level of disturbance. It is of interest, too, that the index also appears independent of whether the child is reported to show severe problem behaviours at school, at home or in both locations. That is, the predominant type of problem behaviours shown by the child does not appear to be systematically associated with the subgroup to which they have been allocated in this study once severity of problem behaviours are controlled.

The distribution of the diagnosis of psychosis in the study group as a whole and across the four subgroups is represented in the following table (Table 14d).
<table>
<thead>
<tr>
<th></th>
<th>moderate problem subgroup n=42</th>
<th>severe school problem subgroup n=37</th>
<th>severe home problem subgroup n=37</th>
<th>severe-pervasive problem subgroup n=43</th>
<th>total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>diagnosed psychotic</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3%</td>
</tr>
<tr>
<td>not diagnosed psychotic</td>
<td>41</td>
<td>35</td>
<td>37</td>
<td>41</td>
<td>97%</td>
</tr>
</tbody>
</table>

**Table 14d.** Distribution of the diagnosis of psychosis across the four subgroups

The overall percentage of children in the study diagnosed psychotic was very small at 3%, even though the Unit was a tertiary referral centre for treating the most severely emotionally and behaviourally disturbed children in the state. This is consistent with the aforementioned recent survey of the 29 inpatient child psychiatric units in the UK (Green & Jacobs, 1998b), where the median percentage of diagnosis on admission was 2.5% for “psychosis”. This reflects the rarity of psychosis in childhood and early adolescence. With only 5 of the 159 cases receiving this diagnosis, this variable – despite its intrinsic clinical interest for the reasons outlined - was dropped as a predictor variable from the study, leaving the externalizing/internalizing index as the only child clinical predictor variable.\(^{58}\)

It is also of interest that the small group of children receiving the diagnosis of psychosis is not significantly clumped among the subgroups. The distribution of the externalizing/internalizing index and the diagnosis of psychosis across the four subgroups suggests that whatever clinically meaningful differences may exist among the subgroups constructed in this study, they appear not to be associated with these two typologies.

\(^{58}\) Other than the parental and teacher total problem scores forming the subgrouping schema.
5.33 Family socio-demographic variables

There are two family socio-demographic variables among the predictors of outcome: socio-economic status and family structure.

The mean and standard deviation of socio-economic status for the four subgroups, and the results of a statistical test of difference across the subgroups (one way ANOVA), are presented in the following table (Table 14e).

<table>
<thead>
<tr>
<th></th>
<th>moderate problem subgroup n=42 mean (s.d)</th>
<th>severe school problem subgroup n=37 mean (s.d)</th>
<th>severe home problem subgroup n=37 mean (s.d)</th>
<th>severe-pervasive problem subgroup n=43 mean (s.d)</th>
<th>P values</th>
</tr>
</thead>
<tbody>
<tr>
<td>family socio-economic status</td>
<td>39.88 (15.71)</td>
<td>46.57 (16.72)</td>
<td>43.73 (14.51)</td>
<td>46.72 (15.08)</td>
<td>.16</td>
</tr>
</tbody>
</table>

Table 14e. Differences in family socio-economic status (Daniel’s scale of prestige) across the four subgroups (one way ANOVA)

These results show no statistically significant differences across the subgroups, with a small trend towards higher socio-economic status (lower scores on Daniel’s scale of prestige) among the moderate problem subgroup. Contrasts between the moderate problem subgroup and the other three subgroups produced $p$ values of .058, .273 and .044, respectively. While there is some suggestion that the moderate problem subgroup may be of higher socio-economic status than the school problem subgroup and the severe-pervasive problem group, the effect is not strong.
The weakness of the effect is perhaps unexpected given the negative relationship between family socio-economic status and child psychiatric morbidity in the general population. For example, the aforementioned Western Australian Child Health Survey found the proportion of children with mental health problems (those with a T score of 60 or above – that is scoring at the 82\textsuperscript{nd} percentile - on the Total Problem Score of either the CBCL or TRF) increased as parental income fell (Zubrick, Silburn, Garton, Burton, Dalby, Carlton, Shepherd & Lawrence, 1995, p. 48). Given that the measure of socio-economic status in this study - family occupational prestige - might be expected to correlate strongly with parental income, and that the measures of psychopathology are the same in both studies, the weakness of the effect of socio-economic status on the subgroupings in the present study needs to be explained.

One possibility is that as suggested with the subgroup differences on gender, this may reflect some aspect of the admission process. Another possibility is that the general relationship in the population between socio-economic status and the presence of moderate child mental health problems is quite different from the relationship between socio-economic status and severity of child mental health problems within the subpopulation of children with more severe psychopathology. To illustrate, a T score of 60 on the total problem score of the CBCL for boys 4 to 11 can be achieved with a raw total problem score of 40\textsuperscript{59}. Referring to the CBCL total problem scores at admission of the children in the present study (Table 8), a score of 40 is approximately one standard deviation below the mean of the moderate problem subgroup. Put another way, the average CBCL total problem score of the severe-pervasive problem subgroup, 102, converts to a T score of 78 for boys aged 4 – 11, and the average CBCL total problem score of the moderate problem subgroup, 52, converts to a T score of 67 for boys aged 4 – 11. (This raises again the question of the meaningfulness of a definition of child mental health

\textsuperscript{59} CBCL/4-18 Profile for Boys – Problem Scales (Achenbach, 1991).
morbidity as those rated at the 82\textsuperscript{nd} percentile of Total Problem Scores; the criteria becomes even more inclusive if a child can be so classified on reaching this threshold on ratings by either their parent or their teacher).

The distribution of the three classifications of family structure in the study group as a whole, across the four subgroups, and the results of a statistical test of differences in distribution across the four subgroups, is represented in the following table (Table 14f).

<table>
<thead>
<tr>
<th></th>
<th>moderate problem subgroup n=42</th>
<th>severe school problem subgroup n=37</th>
<th>severe home problem subgroup n=37</th>
<th>severe-pervasive problem subgroup n=43</th>
<th>total</th>
<th>P values</th>
</tr>
</thead>
<tbody>
<tr>
<td>two natural</td>
<td>23</td>
<td>20</td>
<td>19</td>
<td>19</td>
<td>51%</td>
<td>.787</td>
</tr>
<tr>
<td>step parent</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>9%</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>17</td>
<td>13</td>
<td>13</td>
<td>20</td>
<td>40%</td>
<td></td>
</tr>
</tbody>
</table>

**Table 14f.** Differences in distribution of family structure across the four subgroups: chi square

With regard to the group as a whole, it is difficult to assess how representative it is of the general Australian population owing to the different ways in which family types are classified. For example, the Western Australian Child Health Survey (Zubrick, Silburn, Garton, Burton, Dalby, Carlton, Shepherd & Lawrence, 1995, p.48) reported the following “common living arrangements”: two adults and child(ren) residing together (80%); one adult and child(ren) residing together (19%); more than two adults and child(ren) residing together (1%). It is very difficult to compare this to the present findings as what was felt to be a possibly clinically important distinction between “two natural parent” and “step-parent” families (as outlined, 4.32) was maintained, rather than simply the number of adults co-residing. In any case, it might be expected that the families with children admitted to the Unit are not representative of families in the general population with children showing problem behaviours. An important Unit policy
during the period of the study, especially so in the period previous to 1 June 1996, was only to admit children with an adult or adults strongly committed to intensive involvement with the Unit’s treatment program. This meant a considerable commitment of their time, in (usually) weekly family therapy sessions, attendance at emergency meetings called at short notice following clinically significant events on the Unit, and attendance at other Unit events and meetings with other members of staff (e.g. teachers). This ruled out many families with troubled children at an early stage in the referral process (e.g. when Unit staff conferred with the referring agents), and still others at the assessment stage which took place over weeks or months when family commitment to the work was thereby tested.

With regard to the subgroups, these results show no statistically significant differences in the distribution of family structure across the four subgroups.

It is of interest that whatever clinically meaningful differences may exist among the subgroups constructed in this study, they do not appear to be strongly associated with socio-economic status, and do not appear to be associated with the aspect of family structure examined in the present study.

5.34 Family clinical variable

There was one family clinical variable among the predictors of outcome: parental self-ratings of family functioning. The mean and standard deviation of the general functioning scale of the Family Assessment Device (Version III) for the four subgroups, and the results of a statistical test of difference across the subgroups (one way ANOVA), are presented in the following table (Table 14g).
The subgroups show statistically significant differences with regard to family functioning, which is as expected. The mean parental ratings of family functioning for the moderate problem and the severe school problem subgroups are close to the clinical cut-off of 2.0 for this scale (Miller, Epstein, Bishop & Keitner, 1985, p. 353) while the severe home problem and the severe-pervasive problem subgroups score higher (i.e. indicating more dysfunction). It is not surprising that family functioning is rated to be poorer in the subgroups of children rated to be severely problematic at home. While it is tempting to interpret this finding as confirming the relationship between poor family functioning and child problem behaviours, it needs to be remembered that the ratings are provided in both cases by the parents and that there is significant overlap in the content areas of the General Functioning score of the FAD and the Total Problem Score of the CBCL. This finding, and its ambiguity, is not unexpected given the intrinsic limitations of the FAD and other instruments of this type (as discussed in 4.32).

### Table 14g

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderate problem subgroup n=42</td>
<td>2.05 (.36)</td>
<td>.0001</td>
</tr>
<tr>
<td>Severe school problem subgroup n=37</td>
<td>1.99 (.34)</td>
<td></td>
</tr>
<tr>
<td>Severe home problem subgroup n=37</td>
<td>2.32 (.36)</td>
<td></td>
</tr>
<tr>
<td>Severe-pervasive problem subgroup n=43</td>
<td>2.26 (.36)</td>
<td></td>
</tr>
</tbody>
</table>

The final set of predictor variables to be examined are three treatment variables: the child’s length of admission in the Unit, whether they attended the Junior, Middle or Senior School program, and the proportion of the child’s admission which took place post 1 June 1996.
The mean and standard deviation of length of admission for the four subgroups, and the results of a statistical test of difference across the subgroups (one way ANOVA), are presented in the following table (Table 14h).

The subgroups are significantly different with regard to the length of admission. The mean length of admission is around 8 months for both the moderate problem and the severe home problem subgroups, while it is around 10 months for the severe school problem and the severe-pervasive problem subgroups. That is, length of stay is longer in the subgroups composed of children rated by their pre-admission teacher to show severe problems at school.

This is most likely to reflect the difficulties and delays in returning these children to mainstream schools. The children admitted to the Unit did not always maintain a relationship with their previous school, and there was often a gradual reintegration to a new mainstream school with children being required to demonstrate to the naturally apprehensive school that their at-school behaviour had moderated.

<table>
<thead>
<tr>
<th></th>
<th>moderate problem subgroup n=42</th>
<th>severe school problem subgroup n=37</th>
<th>severe home problem subgroup n=37</th>
<th>severe-pervasive problem subgroup n=43</th>
<th>P values</th>
</tr>
</thead>
<tbody>
<tr>
<td>length of admission</td>
<td>8.35 (4.15)</td>
<td>10.15 (5.06)</td>
<td>8.03 (4.01)</td>
<td>10.72 (5.57)</td>
<td>.0286</td>
</tr>
</tbody>
</table>

Table 14h. Differences in length of admission across the four subgroups (months): one way ANOVA

On the other hand, the children returned to their parents each day (day admission) or each weekend (residential admission). As mentioned, it was an important condition of admission that the adults participate vigorously in the treatment (usually weekly or fortnightly family therapy
sessions). The deliberate policies of (a) only admitting children with parental figures strongly committed to arresting the breakdown in their relationship, where such breakdown is present, (b) maintaining the child’s relationship with his family through only partial admission, and (c) strengthening the child’s place in the family through intensive family therapy, meant that it was rare when length of stay was extended due to a breakdown in the child’s home placement.

The distribution of the children admitted into the Junior, Middle or Senior School program, in the study group as a whole and across the four subgroups, and the results of a statistical test of differences in their distribution across the four subgroups, is represented in the following table (Table 14i).

<table>
<thead>
<tr>
<th></th>
<th>moderate problem subgroup n=42</th>
<th>severe school problem subgroup n=37</th>
<th>severe home problem subgroup n=37</th>
<th>severe-pervasive problem subgroup n=43</th>
<th>total %</th>
<th>P values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junior School program</td>
<td>16</td>
<td>13</td>
<td>12</td>
<td>9</td>
<td>31%</td>
<td>.020</td>
</tr>
<tr>
<td>Middle School program</td>
<td>18</td>
<td>20</td>
<td>21</td>
<td>34</td>
<td>59%</td>
<td>.200</td>
</tr>
<tr>
<td>Senior School program</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>10%</td>
<td></td>
</tr>
</tbody>
</table>

Table 14i. Differences in distribution of the program to which the children were admitted across the four subgroups: chi square

With regard to the group as a whole, a much larger number were admitted to the Middle School program than the Junior School, with the numbers admitted to the Senior School program being particularly small. This reflects the following. First of all, the Middle School program, representing the original residential treatment program, was the largest program in the Unit. Secondly, the Senior School was only being established at the time of commence of data
collection, by the mid-1990s it was already in run down mode, and was finally closed before the completion of the present data collection. Thirdly, as mentioned [4.21], the data collection commenced in the Middle School program with a lag of about 6 months before data collection commenced in the Junior School and Senior School programs. On the whole, differences evident in the study sample with regard to the number of children enrolled in each programs is consistent with admission patterns over the period of the study.

With regard to the subgroups, there is a statistically significant difference with regard to the distribution of programs to which the children were admitted. It is striking that the severe-pervasive problem subgroup contains no participants who were enrolled in the Senior School; this likely reflects the policy of this program, from its inception, to accept much less disturbed individuals than the more established Middle and Junior School programs. The philosophy and orientation of the Senior School program, already reflecting the trend at that time towards a cognitive-behavioural orientation, on one hand, and the need to contain older, larger, physically stronger adolescents, on the other, included the adolescents being admitted on the basis of an explicit, verbal contract with the staff. This contract involved an agreement to curb their dangerous and threatening behaviours.

In contrast, the Middle and Junior School programs, while requiring an explicit contract with the parents, would admit very difficult children with no clear commitment by the children, at least at the beginning, to curb their behaviours. An important philosophy of psychodynamic milieu therapy was that the children’s problem behaviours, including extreme externalizing or acting-out behaviour, represented important communication about relevant developmental needs - the difficult emotions and related relationship issues - underlying their problems. An effort was made to maintain a non-punitive, exploratory attitude to these behaviours. This meant that, for
example, “therapeutic holding” – a way of physically holding a distressed, acting-out child so that both the adult and child could safely “ride-out” the tantrum, while (stamina permitting) the adult reflected out aloud on the child’s distress and the precipitating event in the Unit milieu, often in the context of what was emerging in the course of the admission about family issues and his personal history – was standard practice in the Middle and Junior Schools in the pre June 1996 period.

This approach was, not unreasonably, felt to be impractical with adolescents, and the “holding” of problem behaviours in the Senior School was achieved through verbal reasoning, negotiation and explicit reference to the therapeutic contract.

Another relevant factor was that the Middle School was the only residential unit in the NSW public health system catering to primary school age children, with a long established reputation for admitting very difficult kids. In contrast, the newly established Senior School program did not have a residential component and as there were a range of adolescent services in Sydney and NSW offering similar programs it never came to be seen as “the last resort” in the same way as the Middle school program. It may also be the case that adolescents requiring physical containment were more likely than similar primary school aged children to find themselves in custodial settings within the state’s Juvenile Justice (i.e. corrective services for youths) system.

There are also differences between the subgroups with regard to admission to Junior and Middle School, with the former over represented among the moderate problem subgroup and underrepresented among the severe-pervasive problem subgroup. The Middle School, as the original kernel of the Unit in the 1950s, retained to a greater degree than the Junior School the milieu therapy culture, including the presence of highly experienced and competent male and
female psychiatric nurses with a capacity for the non-punitive emotional containment of very
difficult children, and its orientation towards working with the most disturbed children. The
Junior School, as a later development in the history of the Unit, was almost entirely staffed by
females, including individuals with a personal discomfort or philosophical objection to the more
robust, physically confronting, milieu-therapy approach of the Middle program. This may have
influenced the admission process of the Junior School away from the severe-pervasive problem
subtype more prominent in the Middle School.

The mean and standard deviation of proportion of treatment post-June ’96 for the four
subgroups, and the results of a statistical test of difference across the subgroups (one way
ANOVA), are presented in the following table (Table 14j).

<table>
<thead>
<tr>
<th></th>
<th>moderate problem subgroup n=42</th>
<th>severe school problem subgroup n=37</th>
<th>severe home problem subgroup n=37</th>
<th>severe-pervasive problem subgroup n=43</th>
<th>P values</th>
</tr>
</thead>
<tbody>
<tr>
<td>proportion of treatment post-1 June ‘96</td>
<td>.36 (.47)</td>
<td>.47 (.50)</td>
<td>.48 (.50)</td>
<td>.38 (.46)</td>
<td>.64</td>
</tr>
</tbody>
</table>

Table 14j. Differences in proportion of treatment post-June ’96
across the four subgroups: one way ANOVA

These results show no statistically significant differences across the subgroups. This shows that
despite the change in treatment philosophy taking place in the Unit over the decade of the study,
the types of children admitted and treated did not significantly change, at least according to the
classification variables of problem behaviours rated by parents and problem behaviour rated by
teachers underlying the subgrouping schema.
It is of interest that whatever clinically meaningful differences may exist among the subgroups constructed in this study, they do not appear to be associated with whether the child was admitted in the period before or after June 1996. On the other hand, the differences between the four subgroups with regard to the length of admission and which program they attended appear unremarkable and consistent with Unit policy and functioning during the 1990s.
5.36 *Summary of differences across the subgroups with regard to predictor variables*

The following Table (Table 15) summarises the differences across the subgroups with regard to the nine variables included in the analysis of predictors of outcome.

<table>
<thead>
<tr>
<th></th>
<th>moderate problem subgroup</th>
<th>severe school problem subgroup</th>
<th>severe home problem subgroup</th>
<th>severe-pervasive problem subgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>child socio-demographic</strong></td>
<td>gender</td>
<td>mostly males</td>
<td><em>all</em> males</td>
<td>mostly males</td>
</tr>
<tr>
<td></td>
<td><strong>age</strong></td>
<td></td>
<td></td>
<td>no statistically significant differences</td>
</tr>
<tr>
<td><strong>child clinical</strong></td>
<td>externalizing/ internalizing index</td>
<td></td>
<td></td>
<td>no statistically significant differences</td>
</tr>
<tr>
<td><strong>family socio-demographic</strong></td>
<td>socio-economic status</td>
<td></td>
<td></td>
<td>no statistically significant differences</td>
</tr>
<tr>
<td></td>
<td>family structure</td>
<td></td>
<td></td>
<td>no statistically significant differences</td>
</tr>
<tr>
<td><strong>family clinical</strong></td>
<td>family functioning</td>
<td>near the normal/clinical range cut-off</td>
<td>clinical range</td>
<td></td>
</tr>
<tr>
<td><strong>treatment</strong></td>
<td>length of admission</td>
<td>around 8 months</td>
<td>around 10 months</td>
<td>around 8 months</td>
</tr>
<tr>
<td></td>
<td>Program attended</td>
<td>Middle under represented.</td>
<td>Programs proportionately represented</td>
<td>Middle over represented. Senior unrepresented (n=0)</td>
</tr>
<tr>
<td></td>
<td>proportion of treatment post-June ‘96</td>
<td></td>
<td></td>
<td>no statistically significant differences</td>
</tr>
</tbody>
</table>

Table 15. Summary of differences across the subgroups with regard to predictor variables
Differences between the subgroups are noted on four of the predictor variables.

Ratings of family functioning being worse in subgroups where the ratings of child problem behaviours at home are more severe was not unexpected and likely to be artifactual.

Length of admission being longer in subgroups where the ratings of child problem behaviours at school are more severe was understandable in terms of Unit policy and functioning. The over-representation of attendees of the Middle School program in the severe-pervasive problem subgroup and, conversely, the over-representation of the Junior and Senior School in the moderate problem subgroup were also understandable in this way.

The absence of girls in the severe school problem subgroup, however, is unexpected and not explicable in terms of the readily known aspects of Unit policy and functioning.
5.4 Outcomes

5.4.1 Overall outcome 6 and 12 months after discharge

The outcomes for the study group as a whole show a decrease in Total Problem Score between admission and 6 months post-discharge, which is maintained at 12 months post-discharge (Table 16).

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>standard deviation</th>
<th>minimum</th>
<th>maximum</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>admission</td>
<td>75.56</td>
<td>27.28</td>
<td>24</td>
<td>149</td>
<td>159</td>
</tr>
<tr>
<td>6 months post-discharge</td>
<td>50.72</td>
<td>26.55</td>
<td>8</td>
<td>137.5</td>
<td>133</td>
</tr>
<tr>
<td>12 months post-discharge</td>
<td>49.75</td>
<td>28.30</td>
<td>3</td>
<td>137.5</td>
<td>134</td>
</tr>
</tbody>
</table>

Table 16. Total Behaviour Problem scores (raw) at admission, 6 months post-discharge and 12 months post-discharge

However, the average score at 6 months post-discharge remains in the clinical range according to Achenbach’s criterion. Converted to T-scores, utilizing Achenbach’s table for boys 4-11, the average change can be expressed as a decrease from a T score of 73 to a T score of 66, which remains above Achenbach’s recommended clinical cut-off of T scores of 63, associated with the 90th percentile (Achenbach, 1991, p. 58). Furthermore, while it might be seen to be encouraging that the average total problem score did not deteriorate at 12 months - and the very small change is in the positive direction! - using the same tables the raw score of 49.75 (sd = 28.3) converts to a T score of 66.

---

60 CBCL/4-18 Profile for Boys – Problem Scales (Achenbach, 1991).
Displayed as a boxplot (Fig 4), and including only the 108 cases for whom data is available at each of the three time points, the decrease for the study group as a whole from admission to 6 months post-discharge, and the much smaller change between 6 months post-discharge and 12 months post-discharge, is readily apparent.

Table 17 displays the proportion of children who achieved “clinically significant change” between admission and 6 month post-discharge according to Jacobson’s two-fold criteria. As outlined [2.41], Jacobson’s first criterion is reliable change; that is, determining whether the amount of change in an individual's score is greater than what can be accounted for by the unreliability of the measurement instrument. Using the published test-retest reliability of .93 at one-week for total problem scores, and the standard deviation of 16.4 for total problem scores in the normal population61, a reliable change (ie. improvement or deterioration) on the Total Problem score of the CBCL requires a change of at least 13 points (details of the calculations are

---

61 Average calculated from the tables for the four age/gender subgroups, Appendix B of CBCL/4-18 Profile for Boys – Problem Scales (Achenbach, 1991).
Jacobson's second criterion is whether the subject's post-test score is closer to that of the functional population or that of the dysfunctional population. This is operationally defined as whether the subject has crossed the mid-point between the mean of the dysfunctional population and the mean of the functional population. By this criterion, the cut off point for change in clinical status on the Total Problem score of the CBCL is 40 points.62

| N= 133 | reliable change (≥ 13 points) |
|---|---|---|---|
| clinical status 6 months post-discharge | improved | unchanged | deteriorated |
| “normal” (≤ 40 points) | 42 | 11 | 0 |
| “clinical” (> 40 points) | 48 | 28 | 4 |

Table 17. The number of children showing “clinically significant change” between admission and 6 months post-discharge by Jacobson’s two-fold criteria

A total of 90 children showed reliable improvement (as defined). Of these 42 were functioning at 6 months post-discharge within the “normal” range, while 48 children were still functioning outside the “normal” range. That is, of the 133 children for whom both admission and 6 months post-discharge data is available, only 42 fulfilled Jacobson’s dual criteria of “clinically significant” improvement.

A total of 39 children were “unchanged”. Of these 11 remained in the “normal” range, while 28 remained in the “clinical” range.

While this categorization is intuitively appealing, the limitations of Jacobson’s method, especially when applied uncritically, have been outlined [4.21]. For example, it is tempting to

62 The result is the same whether Achenbach's norms for referred and normal populations in the United States appearing in the manual are used, or the data collected by Sawyer and his group (Sawyer, Baghurst & Mathias, 1992) from referred and non-referred populations in Adelaide are used.
consider the 42 who fulfilled the dual criteria of “clinically significant” improvement as having reached a “gold” standard of clinical change, and the 48 who improved but remained within the “clinical” range as having reached a second-best status, a “silver” standard as it were. However, those in the “normal” range post-discharge are likely to have been closer to the “normal” range, or even within the “normal” range, at admission than those in the “clinical” range post-discharge. It may well be that those children showing the largest amount of change in the total problem score – representing some of the most outstanding clinical results – are to be found among the “silver” outcome group.

To take another example, in a Unit that caters to very disturbed children, the eleven children who remained unchanged and in the “normal” range are likely to include cases where the child showed very specific symptoms that are not well reflected in a general index of problem behaviours such as the Total Problem Score of the CBCL. That is, this score may not be a valid indicator of change in this subgroup.

(This helps to illustrate the basis upon which the data analytic strategy for this study was adopted, such as the subgrouping of children into severe and moderate problem groups. It also helps to illustrate the basis for rejecting methods involving the conversion of the outcome data into categories of “treatment successes” and “treatment failures”, such as logistic regression and discriminant analysis [4.5]. Furthermore, it has been argued that inappropriate methods can systematically obscure the amount of change achieved in the “silver” outcome group [2.51]. As stated, the aim of the present thesis is to focus on therapeutic change achieved among the most disadvantaged youths, those with the most severe and pervasive difficulties - those unlikely to be rated by their parents as functioning in the “normal” group even after achieving large improvement in response to appropriate treatment. The present data analytic strategy aims to
maximise the possibility of detecting variables associated with differences in post-discharge scores within this clinical subgroup – the **severe-pervasive problem subgroup**).

Another interesting result to emerge from this schema is the four children who deteriorated between admission and 6 months post-discharge. Those cases running against the trend towards improvement which appears in the group results confirm the presence of heterogeneity within the study sample with regard to response to treatment [3.63].

Table 18 displays the proportion of children who achieved “clinically significant change” between admission and 12 month post-discharge according to Jacobson’s two-fold criteria

<table>
<thead>
<tr>
<th>clinical status 12 months post-discharge</th>
<th>reliable change (≥ 13 points)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>improved</td>
<td>unchanged</td>
<td>deteriorated</td>
</tr>
<tr>
<td>“normal“ (≤ 40 points)</td>
<td>50</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>“clinical“ (&gt; 40 points)</td>
<td>45</td>
<td>24</td>
<td>6</td>
</tr>
</tbody>
</table>

**Table 18.** The number of children showing “clinically significant change” between admission and 12 months post-discharge by Jacobson’s two-fold criteria

A total of 95 children showed reliable improvement between admission and 12 months post-discharge. Of these 50 were functioning at 12 months post-discharge within the “normal” range, while 45 children were still functioning outside the “normal” range. A total of 33 children were “unchanged”. Of these 9 remained in the “normal” range, while 24 remained in the “clinical” range. Finally, a total of 6 children deteriorated between admission and 12 months post-discharge.
While there are small differences between outcomes at 6 months and 12 months when categorised in this way, it may be useful to exclude possible artefacts which may be associated with differences between the set of 133 subjects for whom both admission and 6 months is available, and the separate set of 134 subjects for whom both admission and 12 months data is available. This is possible by examining the smaller set of 108 subjects for whom data is available at all three time points (Table 19).

| N= 108 |                | reliable change (≥ 13 points) |               |               |
|        | clinical       | improved | unchanged | deteriorated |
|        | status         | 6 months |            |              |
|        | “normal” (≤ 40 points) | 32 | 9 | 0 |
|        | “clinical” (> 40 points) | 37 | 27 | 3 |
|        | clinical       | 12 months |            |              |
|        | status         |         |            |              |
|        | “normal” (≤ 40 points) | 44 | 6 | 0 |
|        | “clinical” (> 40 points) | 36 | 18 | 4 |

Table 19. The number of children showing “clinically significant change” between admission and 6 months post-discharge, and between admission and 12 months post-discharge by Jacobson’s two-fold criteria (restricted to the 108 cases where data is available at all three points in time).

This table (Table 19), which is directly comparable to the box-plot of Fig 4 as they both pertain to the same 108 subjects, shows an overall trend towards more children showing improvement (relative to total problem behaviours at admission) at 12 months post-discharge than at 6 months post-discharge, and more children having moved into the “normal” range at 12 months post-discharge than at 6 months post-discharge. This positive trend appears to be stronger for this restricted set of 108 subjects than when examining the broader sets of 133 subjects for whom admission and 6 months data is available, and the 134 subjects for whom admission and 12
months data is available. The stronger trend in the restricted set of 108 subjects is likely to reflect an artefact of selection\textsuperscript{63}.

Consistent with the stated aims of the present study, the overall outcome 6 and 12 months after discharge data is \textit{not} employed to argue the effectiveness of the treatments [4.1]. It is examined to show that there is (a) evidence of change in the outcome variable between admission and follow-up, (b) evidence of variation in the amount (and direction) of change within the study group, and (c) some evidence to suggest that treatment may be associated with that change. This provides a general justification of the multiple linear regression analysis to follow of the relationship between the selected predictors and the selected outcome. It also provides some justification for the inclusion of treatment variables among the predictors – a study of the interaction between treatment and other variables cannot proceed without evidence to suggest that effective treatment was taking place.

\textit{5.42 Outcomes across the subgroups 6 and 12 months post-discharge}

When the four subgroups are examined separately, each show a decrease in Total Problem Score between admission and 6 months post-discharge, but the results with regard to 12 months post-discharge are not homogenous.

\textsuperscript{63} As discussed [4.4], parents happy to return questionnaires at 12 months post–discharge are likely to differ from those not returning questionnaires with regard to their treatment experience and post-discharge experience, which can be expected to be reflected in the ratings of their child’s behaviour at 6 months post-discharge. This consideration, one among the known potential threats to internal validity, preclude the application of a statistical test to differences between pre-treatment and post-treatment total problem scores and then concluding on the basis of $p$ levels that a “significant” improvement has been found (as discussed [2.41], and pointed out in the studies reviewed in Chapter 3). It bears repeating that while the selection factors are expected to lead to better outcomes in the study sample than in the broader population of children receiving treatment at the Unit between August 1990 and August 1999, it is not
Table 20 shows the means and standard deviations on the outcome scores for each subgroup over the three points of time.

*necessarily* expected that the relationship between predictor variables and outcome is different between the study group and the broader group [4.4].
Table 20. Means and standard deviations of CBCL total problem score at admission, 6 month post-discharge and 12 month post-discharge for the four subgroups

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>admission:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>moderate problem</td>
<td>54.64</td>
<td>13.94</td>
</tr>
<tr>
<td>severe school problem</td>
<td>55.93</td>
<td>13.57</td>
</tr>
<tr>
<td>severe home problem</td>
<td>91.04</td>
<td>15.47</td>
</tr>
<tr>
<td>severe-pervasive problem</td>
<td>102.70</td>
<td>21.49</td>
</tr>
<tr>
<td><strong>F</strong></td>
<td>59.4418</td>
<td></td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td><strong>6 months post-discharge:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>moderate problem</td>
<td>38.86</td>
<td>19.02</td>
</tr>
<tr>
<td>severe school problem</td>
<td>38.65</td>
<td>21.30</td>
</tr>
<tr>
<td>severe home problem</td>
<td>64.90</td>
<td>26.38</td>
</tr>
<tr>
<td>severe-pervasive problem</td>
<td>66.15</td>
<td>30.56</td>
</tr>
<tr>
<td><strong>F</strong></td>
<td>10.3624</td>
<td></td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td><strong>12 months post-discharge:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>moderate problem</td>
<td>28.88</td>
<td>17.21</td>
</tr>
<tr>
<td>severe school problem</td>
<td>39.72</td>
<td>22.16</td>
</tr>
<tr>
<td>severe home problem</td>
<td>55.40</td>
<td>23.34</td>
</tr>
<tr>
<td>severe-pervasive problem</td>
<td>66.15</td>
<td>31.93</td>
</tr>
<tr>
<td><strong>F</strong></td>
<td>12.2498</td>
<td></td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5 presents the results for each subgroup, restricted to the 108 cases (out of the total of 159) where both 6 month post-discharge and 12 month post-discharge outcomes were available, in box plots.
While the *severe school problem* subgroup and the *severe-pervasive problem* subgroup showed little change between 6 month and 12 month outcomes, the *moderate problem* subgroup and the *severe home problem* subgroup continued to show improvement between 6 month and 12 month outcomes. That is, children *without* severe problems at school were reported by their parents to keep improving up to 12 months after discharge, while children *with* severe problems at school were not reported to improve beyond the levels achieved at 6 months post-discharge. This finding held irregardless of whether the children were reported to show severe or only moderate problems at home.

The following four tables (Table 21 - 24) show the number of children in each subgroup who achieved “clinically significant change” between admission and 6 month post-discharge, and between admission and 12 month post-discharge according to Jacobson’s two-fold criteria. They are directly comparable to the box-plot of Fig 5 as they pertain to the same 108 subjects.
Table 21. The number of children in the moderate problem subgroup showing “clinically significant change” by Jacobson’s two-fold criteria (restricted to the 108 cases where data is available at all three points in time).

<table>
<thead>
<tr>
<th></th>
<th>improved</th>
<th>unchanged</th>
<th>deteriorated</th>
</tr>
</thead>
<tbody>
<tr>
<td>clinical status 6mth pst-disch</td>
<td>“normal” (≤ 40)</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>“clinical” (&gt; 40)</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>clinical status 12mth pst-disch</td>
<td>“normal” (≤ 40)</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>“clinical” (&gt; 40)</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 22. The number of children in the severe school problem subgroup showing “clinically significant change” by Jacobson’s two-fold criteria (restricted to the 108 cases where data is available at all three points in time).

<table>
<thead>
<tr>
<th></th>
<th>improved</th>
<th>unchanged</th>
<th>deteriorated</th>
</tr>
</thead>
<tbody>
<tr>
<td>clinical status 6mth pst-disch</td>
<td>“normal” (≤ 40)</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>“clinical” (&gt; 40)</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>clinical status 12mth pst-disch</td>
<td>“normal” (≤ 40)</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>“clinical” (&gt; 40)</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 23. The number of children in the severe home problem subgroup showing “clinically significant change” by Jacobson’s two-fold criteria (restricted to the 108 cases where data is available at all three points in time).

<table>
<thead>
<tr>
<th></th>
<th>improved</th>
<th>unchanged</th>
<th>deteriorated</th>
</tr>
</thead>
<tbody>
<tr>
<td>clinical status 6mth pst-disch</td>
<td>“normal” (≤ 40)</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>“clinical” (&gt; 40)</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>clinical status 12mth pst-disch</td>
<td>“normal” (≤ 40)</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>“clinical” (&gt; 40)</td>
<td>17</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 24. The number of children in the severe-pervasive problem subgroup showing “clinically significant change” by Jacobson’s two-fold criteria (restricted to the 108 cases where data is available at all three points in time).

<table>
<thead>
<tr>
<th></th>
<th>improved</th>
<th>unchanged</th>
<th>deteriorated</th>
</tr>
</thead>
<tbody>
<tr>
<td>clinical status 6mth pst-disch</td>
<td>“normal” (≤ 40)</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>“clinical” (&gt; 40)</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>clinical status 12mth pst-disch</td>
<td>“normal” (≤ 40)</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>“clinical” (&gt; 40)</td>
<td>13</td>
<td>7</td>
</tr>
</tbody>
</table>
These tables show the same finding – that children in the two subgroups *without* severe problems at school were reported by their parents to keep improving up to 12 months after discharge, while the children in the two subgroups *with* severe problems at school were not reported to improve beyond the levels achieved at 6 months post-discharge – but with perhaps less clarity than the box-plot format (Fig 5). For example, while Table 22 shows very little difference in the **severe school problem subgroup** between 6 months post-discharge and 12 months post-discharge on either of the two criteria, in agreement with the graphic format of Fig 5, Table 24 suggests perhaps some small positive difference between 6 months post-discharge and 12 months post-discharge which is not evident in Fig 5. This is probably an artefact of the reduction of the data required by the application of Jacobson’s criteria, providing further justification of reason to rejection of a data analytic strategy involving the reduction of the CBCL outcome data into these, or similar, outcome categories. Moreover, there does not appear to be a systematic association between Jacobson’s outcome categories and the four subgroups beyond what is revealed in the group data presented as box-plots.

On the other hand, the finding of what appears to be a systematic difference between the four subgroups with regard to 6 month and 12 month outcomes provides support for the adopted strategy, and in particular, the clinical meaningfulness of *this* subgrouping.
5.5 Predictors of outcome

5.5.1 Strategy

As stated, we were interested in examining eleven selected variables with regard to their
combined ability to predict 6 month post-discharge outcomes and 12 month post-discharge
outcomes differentially within the four subgroups.

Multiple linear regression analysis is a statistical technique which allows the relationship
between one outcome variable and several predictor variables to be assessed. It is particularly
useful for naturalistic studies, where the predictor variables may well be correlated with one
another and with the outcome variable to varying degrees (Tabachnick & Fidell, 1996). The
procedure results in an equation that best predicts outcome from the selected predictors in the
following form:

\[ \mu_Y = \alpha + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_pX_p \]

where \( \mu_Y \) is the average predicted value of the outcome, \( \alpha \) is the constant, the Xs stand for each
of the predictor variables and the \( \beta \)s stand for the coefficients assigned to each X (i.e. the effect
on \( y \) of a 1-unit increase in X with other Xs held constant).

The important departure from the conventional application of multiple linear regression in the
present study is that the analyses are performed within the four clinical subgroups, so that the \( \beta \)s
are free to vary for each subgroup, to arrive at four equations each potentially quite different in
this respect. That is, a predictor may be important in one equation (i.e. in one clinical subgroup)
and not important in another, and when the equation is reduced to a model which contains only
the most important predictors, different predictors might be retained in different subgroups. In
this way, rather than developing one model of predictors for the subjects in the study, the
development of a different model for each clinical subgroup allows for the possibility of four quite distinct models – based on the same set of eleven predictors – to emerge.

The models fit to the entire data set were of the form:

\[ y = \alpha + \beta_S X_S + \beta_H X_H + \beta_P X_P + \delta_M Y_M + \delta_S Y_S + \delta_H Y_H + \delta_P Y_P \]

where \( X_S = 1 \) for subjects in the **severe school problem** subgroup
\[ X_S = 0 \] for all other subjects

where \( X_H = 1 \) for subjects in the **severe home problem** subgroup
\[ X_H = 0 \] for all other subjects

where \( X_P = 1 \) for subjects in the **severe-pervasive problem** subgroup
\[ X_P = 0 \] for all other subjects

and where for any predictor \( Y \) (e.g. length of admission) we break \( Y \) into 4 sub-variables as follows:

\[ Y_M = Y \] for subjects in the **moderate problem** subgroup
\[ Y_M = 0 \] for all other subjects

\[ Y_S = Y \] for subjects in the **severe school problem** subgroup
\[ Y_S = 0 \] for all other subjects

\[ Y_H = Y \] for subjects in the **severe home problem** subgroup
\[ Y_H = 0 \] for all other subjects

\[ Y_P = Y \] for subjects in the **severe-pervasive problem** subgroup
\[ Y_P = 0 \] for all other subjects

Thus, this model for each subgroup becomes:

**moderate problem** subgroup: \[ y = \alpha + \delta_M Y \]

**severe school problem** subgroup: \[ y = \alpha + \beta_S + \delta_S Y \]

**severe home problem** subgroup: \[ y = \alpha + \beta_H + \delta_H Y \]
severe-pervasive problem subgroup: \[ y = \alpha + \beta_p + \delta_p Y \]

In this way, distinct regressions can be fit in each subgroup as the data requests.

The necessity for distinct regressions can be tested using increment tests against models constrained by \( \delta_M = \delta_S = \delta_H = \delta_p = \delta \) and/or \( \beta_S = \beta_H = \beta_p = 0 \)

The reduction strategy for arriving at a predictive model for each subgroup had two stages. The first stage was as follows.

(a) Examining the full model, that is, the prediction of outcomes utilising all predictor variables in all subgroups.

(b) Commencing with the first subgroup, systematically dropping predictors where the partial \( p \) value from the multiple regression was > .10. The resulting reduced model is then compared with the full model using the increment test. Where the increment test is significant at the .10 \( p \) value, the dropped predictors are reintroduced into the model, commencing with lowest \( p \) value until the increment is no longer significant at the .10 \( p \) value. The relatively conservative \( p \) value of .10 is set for this stage to guard against prematurely dropping variables which might play important roles in the second stage.

(c) This reduction process is then applied, in turn, to the second subgroup (i.e. with the full set of variables from the first subgroup in place). The process is then applied in turn to the third and fourth subgroups, each time utilising for this purpose the full set of variables for the subgroups not undergoing the reduction process.

(d) The reduced set of predictors for each subgroup from steps (b) & (c) are then combined to form the partial model which is then compared to the full model using the increment
test. Where the increment test is significant at the .10 \( p \) value, the dropped predictors are reintroduced into the model, commencing with the variable with the lowest \( p \) value until the increment is no longer significant.

The second stage involved a systematic attempt at further reductions in the partial model (partial model\(_1\)).

(a) Variables with partial \( p \) values \( > .05 \) are dropped, with the reduced model (partial model\(_2\)) being compared against partial model\(_1\) using the increment test, set at the \( p \) value of .10.

(b) Where the increment test is “passed” the reduced model is accepted as the final model. Where the increment test is “failed”, the dropped predictors are reintroduced into the model (partial model\(_3\), partial model\(_4\), and so on), commencing with the variable with the lowest \( p \) value, until the increment is no longer significant.

This two stage procedure is carried out according to parameters of statistical significance (\( p \) values) set to achieve a balance between the risk of including in the model predictors which do not contribute (beyond chance) to the prediction of outcome, and excluding from the model those that do.

Only the final model for each subgroup, for 6 months and 12 months outcomes, will be interpreted below (5.52 and 5.53, respectively). The pattern emerging from these eight models (four subgroups at two points of outcome measurement) will be examined with regard to clinical significance in the following chapter (Chapter 6).
5.52 Predictors of 6 months outcome

The variables and constants of the final model, for each subgroup, of predictors of outcomes 6 months after discharge appear in Table 25. A detailed output for the regression analysis appears in Appendix I.

<table>
<thead>
<tr>
<th></th>
<th>moderate problem subgroup estimate (p-value)</th>
<th>severe school problem subgroup estimate (p-value)</th>
<th>severe home problem subgroup estimate (p-value)</th>
<th>severe-pervasive problem subgroup estimate (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant (all subgroups)</td>
<td>35.3 (0.0000)</td>
<td>35.3 (0.0000)</td>
<td>35.3 (0.0000)</td>
<td>35.3 (0.0000)</td>
</tr>
<tr>
<td>constant (specific to subgroup)</td>
<td></td>
<td></td>
<td>29.62 (0.0000)</td>
<td></td>
</tr>
<tr>
<td>CHILD SOCIO-DEMOGRAPHIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td></td>
<td></td>
<td>4.34 (0.0002)</td>
<td></td>
</tr>
<tr>
<td>CHILD CLINICAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ext/int index</td>
<td></td>
<td></td>
<td></td>
<td>-51.03 (0.0315)</td>
</tr>
<tr>
<td>(range = 0.03 to 0.82, higher = more ext.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAMILY SOCIO-DEMOGRAPHIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s.e.s</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>two natural parents</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 0, Y = 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>step family</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 0, Y = 1)</td>
<td></td>
<td></td>
<td>-20.31 (0.0626)*</td>
<td></td>
</tr>
<tr>
<td>“other” family</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 0, Y = 1)</td>
<td></td>
<td></td>
<td>17.20 (0.0270)</td>
<td></td>
</tr>
<tr>
<td>FAMILY CLINICAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>family functioning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TREATMENT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>length of admission</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junior School</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 0, Y = 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle School</td>
<td></td>
<td></td>
<td></td>
<td>-22.71 (0.0388)</td>
</tr>
<tr>
<td>(N = 0, Y = 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior School</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 0, Y = 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>post-June '96</td>
<td></td>
<td></td>
<td></td>
<td>20.20 (0.0257)</td>
</tr>
<tr>
<td>(none=0, all=1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 25. Predictors of outcome 6 months after discharge.*

(* while the p value is >.05, this variable was retained when dropping it caused the increment test, set at the p value of .10, to be failed).
The model accounts for a large amount of variance in outcome ($R^2 = 0.335$, adjusted $R^2 = 0.292$), satisfying Cohen’s definition of large effect size (set at $r^2 = .25$).

As can be seen, the predictors of outcome are quite different for each subgroup.

There are no predictors for the **moderate problem** subgroup in the final model; the best estimate for outcome is the constant for all subgroups.

There is one predictor for the **severe school problem** subgroup - family structure - with “other” family structures increasing total problem scores at 6 months after discharge.

There is one predictor for the **severe home problem** subgroup - again family structure - but with a step-family structure *decreasing* total problem scores.

The pattern of predictors for the **severe-pervasive problem subgroup** is the most complex, and the model includes child, family and treatment variables. While the child’s age has a negative effect on the outcome (older children have higher total problem scores), treatment in the Middle School program (for primary school children) has a more positive effect than treatment in the Junior School (for infants school children)\(^{64}\). The relationship between the age and program membership is very interesting as they are operating independently in opposing directions. This indicates that the optimal outcomes are achieved for younger children within the Middle (primary school) program. As with the **severe home problem subgroup**, children from step families have a better outcome. Furthermore, children with a predominantly externalising
presentation have a better outcome than children with a predominantly internalising presentation (the range was .03 to .82, with a lower index represents a predominantly internalizing presentation and a higher index representing a predominantly externalizing presentation). Finally, children who underwent treatment after 1 June 1996 have a poorer outcome than those who underwent treatment before 1 June 1996.

5.52 Predictors of 12 months outcome

The model of predictors of outcomes 12 months after discharge accounts for an even larger amount of variance in outcome than the model of predictors of outcomes 6 months after discharge ($R^2 = 0.373$, adjusted $R^2 = 0.353$). The variables and constants for each subgroup appear in Table 26. A detailed output for the regression analysis appears in Appendix J.

There are fewer variables in the models for 12 months outcomes than in the models for 6 months outcomes, with stronger partial $p$ values.

Again, the predictors of outcome are different for each subgroup.

There are no predictors for the moderate problem or the severe school problem subgroups; the best estimate of outcome for both subgroups, the constant for all subgroups, is very similar to that in the model for 6 month outcomes.

---

64 As noted (please see Table 10i), there were no admissions to the Senior School program within this subgroup.
Table 26. Predictors of outcome 12 months after discharge.

There is one predictor from the **severe home problem** subgroup, with longer admissions associated with worse outcomes.

The model of predictors for the **severe-pervasive problem** subgroup is again the most complex, and again includes child, family and treatment variables. The family socio-economic status has a
positive effect on the outcome (children of higher socio-economic status have lower Total Problem Score). Furthermore, in a repeat of findings from the 6 month outcomes, children from step families have better outcomes, and again children who underwent treatment after 1 June 1996 have poorer outcomes than those who underwent treatment before 1 June 1996.

5.54 Summary

The full list of predictor variables in the study, with estimates and \( p \) values for those included in the model for 6 months outcomes and 12 months outcomes, appear in the table below (Table 27). This displays the eight models (4 subgroups x 2 outcomes) in a way conducive to an examination of the relationships between them.

There is little stability between the models for predictors of 6 months outcomes and the models for predictors of 12 months outcomes for the moderate problem, severe school problem and severe home problem subgroups, with no variables appearing in both models for any of these subgroups. However, the 6 months and 12 months outcome models for the severe-pervasive problem subgroup overlap with regard to two variables: membership of step-families and the proportion of treatment taking place after 1 June 1996. Being a member of a step family and treatment prior to 1 June 1996 lead to better outcomes for children in this subgroup at both follow-up periods.

With regard to predictor variables across the eight models, these are only two variables that appear more than once: again, membership of step-families and the proportion of treatment taking place after 1 June 1996. Apart from being included in both the 6 month and 12 month outcome models for the severe-pervasive problem subgroup, as noted, membership of a step
family is also associated with positive 6 month outcomes for the severe home problem subgroup.
Table 27. Predictors of outcome 6 months and 12 months after discharge.
6.1 Introduction to this Chapter

The discussion of results will take place in three sections. The first section pertains to a consideration of pre-treatment differences found across the subgroups [6.2], the second section pertains to a consideration of the outcomes achieved [6.3], and the third section pertains to findings about the predictors of outcomes [6.4].
6.2 Pre-treatment differences across the subgroups

The differences between the subgroups with regard to the predictor variables were mostly unremarkable. As mentioned, they were either expected (parents who rated their child to show severe problem behaviours at home gave higher scores of family dysfunction than parents who rated their of child to show moderate problem behaviours at home) or explicable in terms of the Unit policies and functioning (length of admission was longer in children with severe school problems, who might be expected to require more time to reintegrate to community schools; the uneven distribution of subgroups across the Junior, Middle and Senior schools likely reflects differences in the aims and capacity of the three programs). Furthermore, the similarity across the four subgroups on the externalizing/internalizing index suggested that the subgroups did not differ with respect to this aspect of a child’s presentation.

On the whole, the descriptive data showed few differences across the four subgroups other than those variables forming the basis for the subgrouping.

There was one unexpected result, however. The absence of girls in the severe school problem subgroup remains difficult to explain. The number of girls in the study is not large, and the \( p \) value for the test of statistically significant difference in gender distribution across the subgroups of 0.14 is not compelling. If it does reflect an aspect of the Unit’s policies and functioning, it defies an easy explanation. One possible explanation is considered below [6.32].
6.3 Outcomes

Outcomes are first discussed for the study sample as a whole [6.31] and then in relation to the four clinical subgroups [6.32]

6.31 Outcome for the study sample 6 and 12 months after discharge

As discussed in Chapters 2 and 3, in the absence of a control group a statistical test of difference in mean Total Problem Score from pre-treatment to follow-up is not meaningful. Similarly, other approaches to evaluate effect size and compare them across different studies, and their limitations, have been discussed [3.51]. An alternative approach, to compare raw scores across studies, was put forward as a simple method for allowing clinically significant patterns to emerge (Table 4, Figure 1). These studies provide a context within which the clinical significance of the present results can be addressed.

The “less effective” treatments remain to the left (i.e. those evaluated by Cohen, Bradley & Kolers, 1987, Kiser et al, 1996 and Kotsopoulos et al, 1996), the “more effective” treatments are in the middle (i.e. those evaluated by Grizenko and colleagues, and Kazdin et al, 1987), with the outcomes for the present study added to the right (Fig. 6).

1. The pre-treatment level of Total Problem Behavior Score for the present study is less severe than that reported for the “more effective” treatments, but generally more severe than the “less effective” treatments (although it is difficult to compare directly with the Internalizing and Externalizing Scores of Kotsopoulos et al, 1996).
2. The amount of change seen in the present study from pre-treatment to followup is smaller than that reported for the “more effective” treatments. It is similar to that seen in the one study among the “less effective” treatments which reports followup data (Kiser et al, 1996). The amount of change in the present study can also be compared to that for the control groups (reported by Cohen, Bradley & Kolers, 1987, Grizenko, Papineau & Sayegh, 1993 and Kazdin et al, 1987), and the loss of the initial small gain in the control group by 12 month post-discharge reported by Kazdin et al (1987). These comparisons suggest that the amount of change seen at followup in the present study is not an artefact of the repeated test administration and related procedures, and suggests that genuine therapeutic gains have been achieved and sustained.

3. The post-treatment level of Total Problem Behaviours in the present study remains in the clinical range, as was the case in all treatment groups except those reported by Grizenko and colleagues.

4. The gain at 6 months post-discharge in the present study was maintained at 12 months post-discharge. While this provides further support for the view that genuine therapeutic gains have been achieved, there is no directly comparable data available from the other studies as there was generally only one point of followup. The exception is the study by Grizenko (1997), where the second followup took place 5 years post-discharge.

This comparison suggests that the present study sample, on average, showed improvement 6 months and 12 months post-discharge which was only moderate, but too large to be dismissed as artifactual. Furthermore, on average, the children’s level of problem behaviours at followup still places them in the “clinical range”.
Figure 6. Graphic summary of evaluative studies reporting CBCL Total Problem Scores (except Internalizing and Externalizing scores where indicated) including the present study.
Again, as previously outlined [5.41], it needs to be emphasised that this discussion does not pertain to the general effectiveness of Arndell Child and Adolescent Unit during the period of the study. The importance of demonstrating that genuine therapeutic gains were likely to have taken place in the study sample pertains to the meaningfulness of the main analyses to follow. Without it, the study of predictors of outcome addresses only the child and family characteristics associated with good prognosis, and not the relationship between patient and treatment factors associated with therapeutic gain, which is of main clinical interest.

6.32 Differences in outcome across the subgroups 6 and 12 months post-discharge

The unanticipated finding that children without severe problems at school were reported by their parents to keep improving between 6 months post-discharge and 12 months post-discharge, while children with severe problems at school were not reported to improve beyond the levels achieved at 6 months post-discharge, is very striking.

One possible explanation is that children with severe school-based problems did not receive as much effective assistance with specific school-based problems as those with severe home-based problems received with regard to specific home-based problems. Given the emphasis of the Unit on the child’s emotional and behavioural problems, on one hand, and the difficulty of engaging severely emotionally disturbed and behaviourally disruptive children in academic work, on the other, it may be the case that children with severe school-based problems experienced specific academic difficulties which remained undiagnosed. For example, formal cognitive testing administered on the Unit tended to have a neuropsychological focus, consistent with the clinical specialization of the psychologists, and formal psychological testing of the academic strengths and weaknesses of the children did not routinely take place. The attention to the diagnosis and
treatment of, for example, specific learning difficulties, consistent with the focus of a specialist educational psychologist, was almost entirely absent among the clinicians. Similarly, occupational therapists on the Unit tended to work generically as clinicians and case-managers, in line with other mental health professionals, and were not engaged in the assessment and treatment of dyspraxia or other specific learning difficulties.

Furthermore, it might be expected that any such specific learning difficulties remained unrecognised and untreated upon the child’s return to a mainstream school. Given the requirements of class room teaching, the generalist teacher might be expected to be happy just to contain the disruptive behaviours of these children.

The hypothesis that emotionally disturbed children with severe school behaviour problems show limited improvement in home behaviour problems unless they receive attention to specific academic difficulties, is in principle testable. Yet - as was shown in Chapter 3 to be the case with studies of specialist residential or day care facilities for severely emotionally disturbed children - the literature on the association between academic achievement and psychosocial functioning is vast yet lacking conceptual clarity and methodological rigour (Rourke & Fuerst, 1995). At the same time – as also emphasised in Chapter 3 – enormous pragmatic difficulties need also to be addressed. It needs to be recognised, for example, that the level of difficulty of engaging children with severe school based behavioural problems in (a) a valid assessment of their academic strengths and weaknesses, and (b) appropriate remedial teaching or treatment, are enormous. It likely exceeds that of clinicians in specialist units engaging such children in individual psychotherapy, or that of the nursing staff engaging such children in the day to day requirements of the setting.
Despite these difficulties, the present finding about the difference in home-based outcome between children with and without severe school-based problem behaviours requires attention to be paid to the complex academic and social variables, and their possible interactions, which may be relevant to understanding and effectively treating the latter subgroup. For example, it has emerged in the present literature review [3.41] that the relatively more able students may not show academic improvement or may even show deterioration in academic achievement when attending specialist Units for emotionally disturbed children. The present study neglects the effectiveness of specialist Units in addressing specific academic problems, as opposed to the emotional, behavioural and social problems of the child. Should further research support the hypothesis that undetected and unameliorated academic problems co-existing with emotional/behavioural problems limit the level of positive change in emotional/behavioural problems evident to parents, this will mean that additional, specific interventions for academic problems may be necessary for at least a portion of these children.

Furthermore, it will be recalled that the absence of girls in the severe school problem subgroup remained difficult to explain [6.2]. Prominent among gender differences in psychological functioning thought to have a neurological basis is the greater incidence of learning disabilities in boys (Todd, Swarzenski, Rossi & Visconti, 1995). It may be possible, then, that the association between gender and membership of this subgroup is indicative of an underlying association between untreated learning disability and a distinctive pattern of more severe problem behaviours at school than at home. Alternatively, the association between gender, learning disability and psychosocial adjustment may be less direct. For example, further research may show that the differential role of gender is important in understanding the impact of the experience of academic failure on psychosocial adjustment.
The quite different approaches of the teaching staff at the Unit before and after June 1996 will be discussed in detail below with regard to predictors of outcome [6.42]. It needs to be pointed out, however, that it was not the case that the post-June 1996 period when an educational philosophy emphasising academic goals and on-task class room behaviour was put in place, was generally more effective in reducing problem behaviours at follow-up (as rated by parents) than the pre-June 1996 period when the educational philosophy emphasised the emotional and social needs of the child. It also needs to be pointed out that it was the sub-grouping at the centre of the present data analytic strategy which allowed the potentially clinically important difference in the trajectory of outcome between children with and without severe behaviour problems at school to emerge. A better understanding of what underlies this difference, potentially leading to the development of appropriate treatment strategies, can be expected to lead to a treatment effect specific to the relevant subgroup.
6.4 Predictors of outcome

The data collection upon which the present study is based commenced in 1990 without guidance from reliable empirical findings about predictors of outcome, or even strong theoretical considerations, and certainly without any idea about the radical change in the therapeutic orientation of the Unit to take place during the coming decade. The data analytic strategy adopted is generally unremarkable for a naturalistic study, with the outcome data collected over a 10 year period examined against selected demographic, clinical and treatment variables of interest in multiple linear regression analyses which simply permitted the strongest predictors among them to emerge. The more sophisticated theoretical considerations, which developed during the course of the evaluative research program, lead to the construction of the sub-groupings within which the regression analyses were conducted. While the basis of the sub-groupings was clinically meaningful, and the general notion of conducting regression analyses in sub-groupings was methodologically coherent, there were no indicators prior to performing the analyses that statistically or clinically significant findings would emerge from this particular sub-grouping over any other sub-grouping.

The lack of commonality between the models of predictors which emerged for the four clinical subgroups, especially in view of the large amount of variance accounted for in the final models ($R^2 = 0.34$ for 6 month outcomes, $R^2 = 0.37$ for 12 month outcomes), is interpreted as lending support to the present data analytic approach over the conventional approach (i.e. allowing only one model of predictors to emerge from the entire study sample).

On the other hand, the lack of stability between models for 6 month and 12 month predictors within three of the four clinical subgroups (the moderate problem, severe school problem and severe home problem subgroups) suggests that interpretation needs to proceed with great care.
Consistent with the exploratory aims of the study and the concern throughout this thesis with clinically meaningful results, the interpretive strategy adopted in this discussion is to examine the pattern of results for suggestions of important underlying structure. Accordingly, it needs to be concluded that no clinically important findings have emerged about the predictors of outcome for these three subgroups\(^65\). In concluding that the selected demographic, clinical and treatment variables were not found to be strong predictors of outcome for almost three quarters of the study sample, no implication arises for the effectiveness or otherwise of treatment for these clinical subgroups.

However, two strong predictors of outcome emerged for the **severe-pervasive problem** subgroup: whether the child is a *member of a step-family* and the proportion of a child’s treatment taking place *before or after 1 June 1996*. As noted, none of the adolescents admitted to the Senior School Program were within this subgroup, so the following considerations apply only to the longer established Junior School and Middle School programs (infants and primary school aged children) at Arndell Unit. The severity of problems shown both at home and at school means that the children of the **severe-pervasive problem** subgroup represent the “core target group” for the Junior School and Middle School programs. These are the children least likely during the period of the study to have found other programs in NSW able to manage them. They are the troubled children for whom this joint health-education Unit was originally

\(^65\) The finding of length of admission as a predictor of 12 month outcome in the **severe home problem** subgroups is difficult to understand. One possible explanation which may be worth recording is as follows. The selection criteria for inclusion in this study (i.e. each subject to have at least one parent who completed the CBCL at registration and also at either the 6 month or the 12 month post-discharge follow-up) leads to the exclusion of cases with short admissions due to early breakdown, but not the cases where treatment proves ineffective but becomes prolonged. It may be the case that length of stay serves, for this subgroup, as a proxy for breakdown in a child’s home placement. It would not be surprising if children experiencing more difficulty returning home would be rated by their parents to show a poorer outcome. The main difficulty with this explanation is that length of admission did not emerge as a significant predictor of 6 month outcome for this subgroup. A possible further explanation is that the effect on the child of a breakdown in home placement emerged only at the later time.
designed. Although representing only 27% of the study sample, and very likely a similar proportion of the patient population admitted to treatment during the 1990s, these represent the severely emotionally disturbed children with whom the present thesis is primarily concerned.

The following discussion of findings with regards to predictors of outcome pertains only to this subgroup.

6.41 Step-families

There is very little known about the influence of family structure on outcomes for children undergoing any sort of psycho-social intervention. However, as discussed in Chapter 4 [4.32], a recent Australian study which suggested that there is a correspondence between a child’s level of disturbance and the number of non-father partners to their mother in the home (Najman, Behrens, Anderson, Bor, O’Callaghan & Williams, 1997) strongly influenced the way in which family structure was analysed in the present study. The findings by Najman et al emphasise the negative impact on a child of a “reconstituted” family. This is not very encouraging with respect to the trend among child-bearing couples, observed in Australia and other Western societies over recent decades, towards earlier and more frequent separations and therefore an increase in stepfamily situations (Funder, 1996; Bumpass, Sweet & Martin, 1990).

The present finding, however, might be seen as offering some encouragement in this respect. Within the subgroup of the most disturbed children, those from step-families benefited more than those from two-natural parent or “other” family structures (which included single mothers, single fathers, two adoptive parents, two foster parents, and two grandparents). As stated in Chapter 4, while the mechanisms for the association between membership in a stepfamily and
poor psychosocial functioning in children are not clear, the negative relationship is consistent across different studies. This rules out the possibility that the present finding - that membership of a stepfamily is a predictor of better treatment outcome - is an artefact of an association between these two variables which exists pre-treatment. It would be expected, on this basis, that membership of step-families would be associated with worse outcomes. The present finding indicates that in this subgroup, after controlling for all other predictor variables included in the study, children from stepfamilies have lower Total Problem Scores at follow-up because they responded better to treatment than children from families with two natural parents or “other” family types.

Two possible explanations for this finding will be discussed. First, children from step-families may possess personal qualities which render them more responsive to treatment. Second, step-families may be particularly responsive to this type of treatment. This organisation is for clarity and is not intended to imply that strict distinctions between child and family structure variables can be maintained.

The first explanation requires there to be important clinical differences between children from step-families and children from other family structures associated with their responsiveness to treatment, other than the child variables already controlled in the present study (gender; age; externalizing versus internalizing presentation; severity of problems; and pervasiveness of problem behaviours across home and school settings). As discussed in Chapter 4 [4.32], two hypotheses have been offered as an explanation for the association between child problem behaviours and the number of non-father partners to their mother in the home; on one hand, the direct impact of non-father males entering the child’s home, on the other, the characteristics of mothers who introduce these transitions. While the hypothesis that children from step families

suffer the cumulative effect of either of these two possible negative factors is in principle testable - for example, through measures of the quality of the child’s capacity for relationships (Main & Hesse, 1991; Westen, 1991) which predict for later problem behaviours – it seems unlikely that this type of hypothesised disadvantage is associated with a better response to treatment relative to other children. If these children possess personal qualities which render them more responsive to treatment, it will be necessary to consider differences between children in stepfamilies and those from other family structures as a “protective” or “resilient” factor (Cicchetti and Cohen, 1995). This might be an example of “patterns normally predictive of disorder but which for reasons to be discovered, do not do so with a particular subgroup of subjects” (Sroufe & Rutter, 1984, quoted by Cicchetti and Cohen, 1995, p. 12). It might be hypothesised, for example, that exposure to different adults in the household provide children with a capacity to make better use of adults in the treatment setting.

The arguments for the second explanation of the results, that step-families may be particularly responsive to this type of treatment, can be stated more specifically. The treatment may be particularly suited for families undergoing the specific transition to stepfamilies. Working on the family issues which characteristically emerge at this point while the child is in a robust setting where his problem behaviours can be managed and emotional reactions made sense of may have very specific advantages. These family issues might include, for example, the structural issues of authority, boundaries and roles, and the relationship issues of working alliance between the couple in dealing with the child’s emotional needs and demands, establishing a workable relationship between the child and the step parent where there may not be one, and perhaps even establishing a parenting alliance between the child’s separated natural parents despite the breakdown in the marital relationship.
It needs to be remembered, however, that the step families included in the severe-pervasive problem subgroup of the present study represents a further subset of step families with children with severe problem behaviours both at home and at school. These are stepfamilies with a clear commitment by both adults to intensive work on family issues as well as managing the child’s extreme and pervasive problem behaviours. The quality of the relationship between parent and step-parent may be quite different from that found in the community groups included in the epidemiological studies, cited above [4.32], leading to the association which has been discussed.

The foregoing considerations raise another issue. The better outcomes for these children may be associated with a more acute presentation (i.e. problem behaviours associated with recent family changes) as opposed to chronic behaviour problems. Chronicity - the duration of behaviour problems - was not included among the predictor variables. This represents a limitation of the present study and indicates an area for further research.

6.42 Admission to the Unit before or after 1 June 1996

The finding within the severe-pervasive problem sub-group of a more positive impact of treatment on children from stepfamilies is of clinical and theoretical interest, especially with regard to appropriate treatment for child problems which may be associated with the present trend towards children experiencing cohabitation with an adult with whom they have no biological ties. However, it is restricted to a minority socio-demographic group (9% of the total study sample, and 9% of the severe-pervasive problem sub-group, please see Table 14f). The finding of better outcomes within the severe-pervasive problem sub-group for those undergoing treatment before 1 June 1996 than those undergoing treatment after 1 June 1996 is of more general interest as it is a treatment variable applying irregardless of the child and
family’s pre-treatment demographic or clinical variables. As a treatment variable, it potentially provides information about what can be done to improve outcomes for children of this subgroup.

As stated in Chapter 4 [4.33], the date of 1 June 1996 was chosen for the present study as the point of demarcation between the “old” and the “new” institutional eras. As mentioned, the change in treatment philosophy at the Unit two thirds of the way through the nearly decade long course of the present study was not anticipated at the commencement of the evaluative research program, and offered a “natural experiment” [3.62]. The greater effectiveness of the traditional “milieu therapy” in comparison with the “empirically-based treatment” which replaced is statistically significant, as shown. With differences in outcome between the two institutional eras of 20.2 in CBCL Total Problem Score at 6 months post-discharge and 24.42 at 12 months post-discharge, the effect size is also large and important.

This can be readily seen when placed against the background of other relevant studies reporting results on the Total Problem Score of the Child Behaviour Checklist, examined in Chapter 3 and again in Chapter 6.

The means, standard deviations, minimum and maximum of (raw) CBCL Total Behaviour Problem scores for those of the severe-pervasive subgroup whose period in admission was entirely or substantially (more than half of time in treatment) prior to 1 June 1996, and for those whose period in admission was entirely or substantially (at least half of time in treatment) after 1 June 1996, is presented in Table 28.
<table>
<thead>
<tr>
<th></th>
<th>mean (standard deviation)</th>
<th>minimum</th>
<th>maximum</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>admission</td>
<td>100.6 (22.8)</td>
<td>106.3 (19.5)</td>
<td>74.5</td>
<td>80.0</td>
</tr>
<tr>
<td>6 months post-discharge</td>
<td>55.5 (32.0)</td>
<td>76.3 (21.7)</td>
<td>8.0</td>
<td>43.5</td>
</tr>
<tr>
<td>12 months post-discharge</td>
<td>61.7 (32.5)</td>
<td>83.8 (21.3)</td>
<td>13.0</td>
<td>29.5</td>
</tr>
</tbody>
</table>

Table 28. Total Behaviour Problem scores (raw) at admission, 6 months post-discharge and 12 months post-discharge of those of the severe-pervasive subgroup whose period in admission was prior to 1 June 1996 and those whose period in admission was after 1 June 1996.

Figure 7 places these results, converted into T scores\(^{66}\), against those of previous studies in the format previously seen in Chapter 3 (Fig 1) and Chapter 6 (Fig. 5). For ease of comparison, only those which provide follow up data and only the results from the treatment groups are included. This leaves three treatment evaluations available for comparison against the present results: the evaluation by Kiser et al (1996) described in 3.24; the evaluation by Grizenko, Papineau & Sayegh (1993) described in 3.23; and the experimental study by Kazdin et al (1987) described in 3.43.

The following comparisons can be made between the results of the pre-June 1996 admissions of the present severe-pervasive subgroup (to be called pre-June 1996 admissions for briefness) and the post-June 1996 admissions of the severe-pervasive subgroup (to be called post-June 1996 admissions).

Figure 7. Comparison of the results for those in the severe-pervasive subgroup whose treatment took place before June 1996 and those whose treatment took place after June 1996, against other selected studies (CBCL Total Problem Scores, T scores).
1. The pre-treatment level of problem behaviours of both the **pre-June 1996 admissions** and the **post-June 1996 admissions** are similar to those of what has previously been referred to as the more effective treatments (evaluated by Grizenko and colleagues and Kazdin et al, 1987). They are higher than the pre-treatment level of problem behaviours reported in Kiser et al (1996). This likely reflects the fact that the latter is an unselected group. Those included in the **pre-June 1996 admissions** and **post-June 1996 admissions** were selected partly on the basis of their higher pre-treatment Total Problem Scores on the CBCL, and the group evaluated by Kazdin et al (1987) was selected partly on the basis of high Aggression or Delinquency scores on the CBCL. The very high pre-treatment scores of the groups evaluated by Grizenko and colleagues apparently reflect the population treated at that unit.

2. The amount of change, in T-score units, from pre-treatment to 12 months followup in the **pre-June 1996 admissions** is comparable to the treatment evaluated by Kazdin et al (1987) and Kiser et al (1996), with all three groups showing much smaller change than that seen from pre-treatment to 6 months followup in the treatment evaluated by Grizenko, Papineau & Sayegh (1993)67. On the other hand, the amount of change from pre-treatment to 12 months followup in the **post-June 1996 admissions** is smaller.

3. The Total Problem Behaviour Score at follow up of both the **pre-June 1996 admissions** and the **post-June 1996 admissions** remain in the clinical range (> T score of 63), as is the case for the treatments evaluated by Kazdin et al (1987) and Kiser et al (1996). Only the treatment evaluated by Grizenko, Papineau & Sayegh (1993) shows a Total Problem Behaviour Score at followup in the “borderline” clinical (T scores of 60 to 63) range.

---

67 The use of T-score units obscures differences in the amount of change in the underlying raw Total Problem Score. For example, the lower pre-treatment problem behaviour scores reported by Kiser et al (1996) means that change in T-score units equivalent to the other treatments will reflect a smaller change in raw Total Problem Score units.
In summary, the results of the present study with regard to average outcomes for the pre-June 1996 admissions and the post-June 1996 admissions of the severe-pervasive subgroup are as follows: parental reports indicate that relative to pre-treatment levels of problem behaviours, both pre-June 1996 and post-June 1996 admissions show improvement 6 months and 12 months post-discharge; the level of improvement in the pre-June 1996 admissions is quite substantial while that of the post-June 1996 admissions is less so, although too substantial to be dismissed as an artefact; however, the children’s level of remaining problem behaviours places both the pre-June 1996 admissions and the post-June 1996 admissions in the clinical range.

While, as discussed [3.41], a limitation of these outcomes is their vulnerability to the artefacts of attrition or selection, where the children with the less successfully treatment outcomes are more likely to be lost to follow-up, there is no reason to suggest that the effects of attrition would systematically vary between pre-June 1996 and post-June 1996 admissions. As for the possibility that the children admitted after June 1996 were more difficult to treat or had poorer prognoses than those admitted before June 1996, the regression models (Table 17) suggest that for the 12 months outcomes the raw results (that is, outcomes unadjusted for the effects of the other predictors, which appear in Table 19 and is represented in Fig 6) underestimate to a small degree the differences between admissions pre-June 1996 and post-June 1996.

The results clearly indicate that for the most severely disturbed children admitted to Arndell Unit in the 1990s - the infant and primary school aged children showing severe behaviour

---

68 The differences in Total Behaviour Problem Scores at 6 months post-discharge between the pre-June 1996 admissions and the post-June 1996 admissions is 20.8, and at 12 months post-discharge the raw difference is 22.1 (Table 18). This compares to the $\beta$ values for the variable of “proportion of treatment post-June 1996” (Table 16) in the final regression models for the severe-pervasive problem subgroup of 20.2 for 6 months outcomes and 24.42 for 12 months outcomes.
problems both at school and at home - treatment prior to June 1996 was more effective than treatment after June 1996.

These results warrant, then, an examination of the differences between the two treatment eras. These will be described in relation to the approach to the child, the general expectations about the child’s behaviour, and the approach to Unit functioning.

The approach to the child: With regard to the approach to the child, the primary focus of “milieu therapy” was on the child’s emotions. As stated by Wood & Long (1991), “One of the most important steps in helping troubled students become more realistic and responsible for their behaviour is to help them understand the feelings that drive their behaviour” (p. 6). This involved a very active technique of directly addressing aggressive and out of control behaviour at the time it was taking place. This meant, for example, taking a child through a period of escalated emotion and behaviour - physically holding the child when necessary - until the outburst subsided and the child could show that he had regained his internal controls and was ready to have the external controls removed. Accompanying this was verbal communication by the staff member which moved progressively from identifying the triggers to the incident, identifying the child’s emotions, perhaps also identifying relevant aspects of known personal history which help make sense of the connection between the trigger and the overwhelming emotions, developing a practical solution (which may involve, for example, reparation of some sort for any damage he may have caused).

---

69 This book was a defacto standard text at Arndell Unit in the early 1990s following a period of intensive training for all health and education staff in the method of “Life Space Intervention” by Nicholas Long. Long had participated in the pioneering work in residential treatment for aggressive boys with Fritz Redl in the 1950s, and represented the continuity between the active techniques and treatment philosophies developed and practiced at the earlier time in the US and Arndell Unit (until 1996).
and the child then rejoining the activity. At any time at any place within the milieu – in the classroom, in the designated therapy rooms, in the residence, and on the grounds - there could be one or more extremely distressed child being physically held by a teacher, nurse or clinician while they “ride through” the tantrum together. This was emotionally and physically very demanding on the staff member. It needs to be noted that therapeutic holding was not without its critics even during this era. It was cogently argued by a small minority of clinical staff that therapeutic holding itself could be experienced by the child as traumatic, contrary to the aim of the milieu environment to provide emotional support for children and protection from trauma.

In contrast, the primary focus of the empirically-validated treatment era was on cognitions. There was no value placed on physical holding – and the associated intensity of emotional and physical intimacy - as an important ingredient of treatment in itself. This meant that crises were more likely to be avoided, and where this was not possible, cognitive interventions were commenced after the period of emotional and behavioural escalation passed, when the child was more amenable to conscious reflection and problem-solving. In contrast with the “on-demand” interventions of the milieu therapy era, there was a preference for the cognitive social problem-solving training which would take place in specific groups at set times, with conditions put in place for clear, conscious, cognitive engagement with the task.

The contrast between the two treatment philosophies was most dramatic in the classrooms. In the earlier period the role of the teachers and teacher-aides were not
distinctly different from the role of nurses or of clinical staff, and the teaching of an academic curriculum was very much a secondary consideration taking up perhaps the first two hours of a day. In the later treatment era, the role of the teachers and teacher aides was clearly centred on formal academic teaching and on shaping the student towards appropriate classroom behaviour; the therapeutic role was specifically rejected. For example, in the earlier era if a student came into class of a morning obviously distressed, the teacher would want to know why; they would be comfortable exploring family issues, and would be familiar with the student’s personal history. In the later era, the teacher would acknowledge the student’s distress but encourage him to direct his focus on school work and expect him to behave in a way conducive to academic tasks for himself and the other members of his class.

Furthermore, there were quite different attitudes in the two eras with regard to the use of medication. There was an expectation in the earlier era that as the “hold” of the milieu on the child became established the use of medication would be reduced. This contrasts with the later era where, in line with the current orthodoxy of “multi-modal” psychiatric treatment, there was an expectation that for many children showing problem behaviours there is an optimal combination of pharmacological and psycho-social intervention.

The general expectations about the child’s behaviour: The tenets of “an emotionally healthy environment” basic to conducting milieu therapy, as specified by Redl & Wiseman (1957) and broadly adopted throughout the milieu therapy era, included “some interpersonal gratifications that are not contingent on behaviour”; “a degree
of tolerance for deviancy and regression”; and “gratifying play free from goals associated with highly competitive games, arts and crafts, or other gadgets that require a level of performance that is difficult to achieve” (p.37). This meant that a greater level of swearing, physical testing of staff limits, immaturity, free play, and general chaos was tolerated - if not actually welcomed - in the earlier era than was the case in the later era. For example, it was a commonplace event in the earlier era for boys to climb onto the roofs of the Unit’s buildings and hurl abuse upon the adults below; it was considered, often with affection, as a barometer of the level of distress in the Unit. Within a few months of the arrival of the Unit director in June 1996 marking the commencement of the “empirical treatment” era, who made it quite clear this behaviour was unacceptable, it was shown that it was possible to run the Unit at a high level of distress without having any boys up on the roof. By the late 90s the long tradition of “boys on the roof” had almost completely disappeared.

Another example which may help to characterise the differences between the treatment philosophies was that in the milieu therapy era new staff would be subjected to an often very difficult period of acclimatisation upon arrival at the Unit, when they would be subjected to very harsh and confronting verbal abuse from children from the outset. An obscenity was the universal response to initial attempts to engage the children, who had a way of quickly discerning the novice’s vulnerabilities and setting upon it with cruel zeal. In the later era, there was a very distinct shift in attitude with respect to expectations of the children’s behaviour, as it was considered counter-productive for the children to engage in behaviours which would not be tolerated outside the Unit. This change was, again, most clearly evident in the school setting, where it became expected that the children would
behave in a way consistent with the mainstream schools from which they were referred, and to which they would be returning.

The approach to Unit functioning: This perhaps represented the most radical difference between the two treatment eras. The notion of a hospital as a system composed of interacting subsystems, with the functioning of each subsystem influencing the functioning of other subsystems as well as the system as a whole, and the hospital itself being a subsystem within the broader systems of the health services, and it in turn a subsystem of the wider world – a precursor to systems thinking - is central to the idea of a therapeutic community (Main, 1989). For Main, “the originator of the concept of the therapeutic community”, the health of each system, including individual patients, required it to “monitor, explore and modify its own function” in relation to the other systems – it is this “discursive activity” which is “therapeutic” (Rayner, 1989, emphasis added).

This well known anecdote from Main’s period as director of the Cassel Hospital illustrates this therapeutic principle.

“In the self-scrutiny of every system, anything might emerge as of significance. For instance, it was noticed that a stray cat had turned up at the hospital and settled in. “Why?” said Dr Main. “Because we have a new rash of mice,” said a cook and a porter. “Where do they frequent?” An investigation showed that it was at the back of the hospital, where some patients were prone to throw bits of food out of several windows. It was then found these were the bedroom windows of the patients of a particular medical therapist. This was
too much of a coincidence. Discontent about the therapy was brought out into the open and the therapist worked on this. The food-throwing stopped, the mice disappeared and the cat moved on.” (Rayner, 1989, p. xix).

Such an open way of working is difficult to sustain, but while it is sustained makes available for productive work the uncomfortable “common sense” typically marginalised from the mainstream of thought and work in a system, and into gossip and “off the record” discussions - in Goffman’s (1961) terminology, the institution's “underlife”.

Post-June 1996, the traditional discursive character of the Unit, in particular the opportunities for reflection in whole Unit meetings (involving all educational, health, and administrative staff) and “team” meetings, were replaced with meetings taking place within disciplinary lines conducted in a hierarchical manner familiar in general medicine. For example, the change in the way of working was clearly evident in “team” meetings coming to an end. At these meetings, all the clinicians, nursing staff and education staff working with children in a particular class would discuss each individual child, as well as the dynamics of the classroom, from the perspective of their own experiences as well as the subsystems to which they may have special access. This included, for example, the clinician’s access to the child’s family system, the teacher’s access to the classroom subsystem, in addition to the general discussion of intra-team, inter-team and broader Unit issues.

From the perspective of the later era, meetings taking place within disciplinary groupings and in a formal hierarchical manner was better suited to maintaining focus
on the data of interest – the behaviour of the child – and a more efficient use of staff
time. In this way, there was a very distinct difference between the two eras with
regard to what was considered signal and what was noise; there was a fundamental
shift in the figure-ground configuration.

There are alternatives which must be considered, however, to the conclusion that the superiority
of outcomes achieved pre-June 1996 in the severe-pervasive subgroup indicates that milieu
therapy was more effective than the cognitive-behavioural, “empirically-validated” therapy
which replaced it. A naturalistic study does not offer the same control as an experiment. While
the present data analysis adjusted for the impact on outcome of differences between the two
eras, if any, with regard to the severity of pre-treatment child problem behaviours at home and at
school, the relative dominance of internalizing or externalizing styles of presentation, and
children’s age and gender, and with regard to family structure, socio-economic status and family
functioning, and also with regard to length of admission and the school program to which they
were allocated, it needs to be considered whether there were other factors which influenced the
effectiveness of treatment over the decade.

An alternative explanation for this result which needs to be explored, is that the conditions for
treatment were more favourable in the period from July 1990 to June 1996 than the period from
June 1996 to August 1999 in some way, or ways, other than the differences already outlined
between “milieu therapy” and “empirically-based therapy”. As discussed [2.32], institutions for
the treatment of severely emotionally disturbed children are often volatile and subject to
controversy, scandals, changes in Unit directors, periods of low staff morale, and so forth. These
fluctuations in institutional functioning may permeate across any simple demarcation of a
decade into two distinct eras.
The 1990s was indeed a period of many controversies and change of senior personnel. The event which most influenced the decade, however, took place in 1987. It was discovered at this time that a generally well-regarded, long-term member of the nursing staff was sexually abusing various children on the Unit while on night duty, apparently undetected, for a period of 10 months. As is generally the case with paedophiles, these children were carefully selected. For example, children undergoing treatment for a previous history of sexual abuse were targeted because not only were they more vulnerable to confusions about affection and physical contact, any symptoms of current abuse, such as sexualised behaviour or verbalisation, would be more likely to be interpreted by staff as the activation of previous trauma in the ordinary course of treatment.

The revelation of this abuse was traumatic for the staff, administrators and clinical supervisors closely associated with the Unit during this period. The scarcely believable horror of what had been perpetrated in their midst, undetected for such a long time, and by a trusted and otherwise unremarkable colleague, made untenable the Unit’s perspective of itself as the protector and healer of children, with the abusers of children located “outside”.

Many institutional events of the 1990s cannot be understood without reference to the earlier traumatic events. These include the concern with “a super safe” environment to protect children from “contamination” by the sexually acting-out behaviour of abused children (including those children who were abused on the Unit in 1986-1987); the demotion of individual therapy (a private and secluded event) as the primary intervention and the promotion of family therapy (a necessarily public event); the very bitter conflict between external clinical supervisors and internal Unit management with regard to the location of blame and shame for not having
detected the abuse earlier (paralleling the dynamics of families when sexual abuse taking place within it is revealed); and the attempt at the close of 1990s to distance the Unit from its past by adopting a new name, “Coral Tree Family Service”. The later era also saw the disposal of the accumulated store of “family” memories contained in the numerous photographs, art works, and other records of the children from the earlier era.70

The following table (Table 29) provides a time line of some of these institutional events, commencing with the sexual abuse of 1986 - 1987 and ending in September 2001 when the evaluative research program was terminated. These events are taken from the minutes of the weekly Unit meetings. Not unexpectedly, these minutes do not directly record the most significant institutional events, as they remained “off the record”. For example, the minutes of July 1995 indicate that a “review” of the Unit is announced, and the minutes of meetings around this time indicate the Head of Department being on “sick leave”, and later on “study leave”. It does not indicate the tensions at this time with the Head of Department threatening disciplinary actions against his own Unit for failure to comply with his directions, and staff concerns about his increasingly erratic behaviour, prior to his departure from the Unit.

70 The termination of the evaluative research program, and later, the destruction of the accumulated protocols and records, may also reflect the same impulse within the institution to distance itself from this traumatic event. While the former represents a legitimate management action, the latter – by individuals not lacking a sophisticated understanding of the Unit’s obligations to the University, under whose control the research was conducted, as well as its obligations to RADGAC and the funding bodies – does not. It suggests the activation of intense emotional forces in the institution. Some hypotheses with regard to unarticulated anxieties associated with fear of humiliation and retaliation were outlined in 2.33.
<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 1986</td>
<td>The sexual abuse of children in the Unit by a nurse on night duty</td>
</tr>
<tr>
<td>February 1987</td>
<td></td>
</tr>
<tr>
<td>November 1987</td>
<td>Disclosure by children of the sexual abuse</td>
</tr>
<tr>
<td>May 1990</td>
<td>Inquiry into events of ’86-’87 commences; Unit Director/ Head of Department (A) announces departure</td>
</tr>
<tr>
<td>July-Aug 1990</td>
<td>Senior School commences. AERP commences; B is made Unit Director, C is made Dept. Head</td>
</tr>
<tr>
<td>October 1990</td>
<td>Meetings re architectural changes to dormitory to ensure “super-safe” environment.</td>
</tr>
<tr>
<td>December 1991</td>
<td>B announces 12 months leave from Unit; D to be acting Unit Director</td>
</tr>
<tr>
<td>December 1992</td>
<td>Announcement that D is to remain as Unit Director</td>
</tr>
<tr>
<td>February 1993</td>
<td>B returns as member of staff</td>
</tr>
<tr>
<td>September 1994</td>
<td>B announces resignation</td>
</tr>
<tr>
<td>June 1995</td>
<td>Announcement that C is on sick leave; D is acting Head of Department (HOD) for “6 months”</td>
</tr>
<tr>
<td>July 1995</td>
<td>“Review” is announced.</td>
</tr>
<tr>
<td>November 1995</td>
<td>“Review” results in “39 recommendations” including “no more inquiries”</td>
</tr>
<tr>
<td>February 1996</td>
<td>C given leave for indefinite period</td>
</tr>
<tr>
<td>June 1996</td>
<td>D (Unit Director and acting HOD) announces departure; E commences as Unit Director</td>
</tr>
<tr>
<td>November 1996</td>
<td>Announcement of the closure of the Senior School</td>
</tr>
<tr>
<td>February 1998</td>
<td>E resigns as Unit Director</td>
</tr>
<tr>
<td>August 1998</td>
<td>Another “review” commences</td>
</tr>
<tr>
<td>February 1999</td>
<td>F commences at Unit</td>
</tr>
<tr>
<td>June 1999</td>
<td>Unit Director position advertised</td>
</tr>
<tr>
<td>February 2000</td>
<td>F appointed Unit Director</td>
</tr>
<tr>
<td>June 2001</td>
<td>New sign “Coral Tree Family Service” is put up</td>
</tr>
<tr>
<td>September 2001</td>
<td>Arndell Evaluative Research Program (AERP) terminated</td>
</tr>
</tbody>
</table>

**Table 29.** Time line of institutional events April 1986 – September 2001
Despite these limitations, the timeline does indicate that the overt crises in the Unit, including the rapid turnover of incumbents in the positions of Unit Director and Head of Department, do not favour the period July 1990 to June 1996 over the period from June 1996 to August 1999. On the contrary, it could be argued that staff was exposed to more crises and instabilities in management in the earlier period than the latter period. As for the level of experience of the Unit directors and Heads of Department, the incumbents during the earlier period (B, C, D) were junior child psychiatrists with no prior management experience (all completed their training on the Unit under A, who was director during the period that the abuse was discovered), while the incumbents during the later period (E, F) were senior child psychiatrists with very substantial previous experience managing child psychiatric units. Furthermore, it has already been mentioned that changed expectations meant that the overt behaviour of the children was much less chaotic, verbally abusive and limit-testing in the later period than the earlier period.

Therefore, it could be argued that milieu treatment was more effective than the empirically-validated cognitive-behavioural treatment *despite* the crises, less experienced management, and overtly chaotic behaviour by the children which were a feature of the earlier period.
CHAPTER 7.
CONCLUSIONS

Two questions underlie this thesis: what accounts for differences in outcome among severely emotionally disturbed children undergoing intensive treatment, and how does one find out.

The latter question has led to the strong emphasis on identifying methodological, conceptual and pragmatic issues impeding progress in this area, and then proceeding to develop methods which adequately address the question of interest, the phenomena of interest, the nature of the data, and the terrain in which this data is embedded.

The former question was addressed through the examination of the relationship between a set of predictor variables and a selected outcome variable in children admitted to a particular treatment centre over a particular period of time. Answers were able to be provided for one quarter of the study sample - forty three children of infant and primary school age who showed severe problem behaviours both at home and at school, admitted to Arndell Child and Adolescent Unit between 1990 and 1999.

It was found that those receiving treatment during the earlier “milieu treatment” era had better outcomes than those receiving treatment during the “empirically-validated treatment” era which superseded it. It was also found for this subgroup that better outcomes were achieved with those from stepfamilies than those from two natural parents or other family structures. While it might be said that these findings pertain only to a subgroup of the study sample, this subgroup represents the most severely disturbed children treated during this period of time. The subgroup appears comparable, on the basis of the Total Problem Score of the CBCL, to the study sample
of other relevant studies which have been discussed in detail (Kazdin et al, 1987; Grizenko, Papineau & Sayegh, 1993).

It is an important finding that this extreme subgroup of children and adolescents achieved better outcomes, as defined, under one type of treatment than another type of treatment. As to whether this finding supports one “brand” of treatment over another, the limitation of these kinds of distinctions have been discussed [3.51]. It has been noted, for example, that the model of treatment for externalizing youth branded cognitive-behavioural by Kazdin et al (1987) has much in common with the psychoanalytic intervention method developed by Redl (Redl & Wineman, 1951, 1952). It is not claimed that the present results justify general conclusions with regard to the greater effectiveness of milieu therapy, described according to psychodynamic and group dynamic principles (Aichorn, 1935; Redl & Wineman, 1951, 1952; Trieschman, Whittaker & Brendtro, 1969) over “empirically-validated” treatments for children and families, described according to behavioural and cognitive-behavioural principles (Kazdin, Esveldt-Dawson, French & Unis, 1987; Kazdin, Siegal & Bass, 1992).

The present results, however, require that the distinct differences between psychodynamic milieu therapy and the cognitive-behavioural approaches, as outlined, are not prematurely excluded from the critical thinking which still needs to take place about the “active ingredients” of successful treatment, at least for the subgroup described as “severe-pervasive” in the present study. The results do not support the view that treatments have become universally more effective under the requirement which emerged during the 1990s that only treatment types already shown to be effective in randomised controlled trials should be carried out in the field. The present results indicate, in fact, that under this initiative a more effective treatment was replaced with a less effective treatment. It is a reminder that any categorisation of treatment, at
least for this population, into “empirically validated” and “not empirically validated” is premature. It also illustrates the problem, discussed above [2.3], of the call for evidence and evidence-based practice serving as a defence against the collection and examination of evidence.

It has been argued that the sorts of methodological and conceptual issues surveyed in this thesis are not “nit picking” issues which clinicians can safely leave to others more interested in research method. If this is true, the present results also suggest that for clinicians the process of critical thinking about the active ingredients of effective treatment is aided by relevant data about the therapeutic enterprise in which one is actually engaged. This is discussed further, below, in relation to the unexpected findings which raised questions about the way in which educational needs have been met in this setting.

There remain many empirical questions, such as the cost-effectiveness of the treatments. It might be asked whether the greater effectiveness of the milieu treatment is justified by what may be its greater costs. The neglect of the costs of treatment is a limitation of the present study, a limitation found in all the studies reviewed. The present results justify a further study of the costs of treatment over the decade, the raw data for which (staff salaries and so forth between 1990 and 1999) is likely to be available.

There are also considerations of the costs of treatment beyond the monetary. A comparison has been made, for example, between the much greater level of acting-out and emotional expression by the youths which was tolerated - and, in fact, encouraged - by staff during the “milieu therapy” era and the more strict limits imposed on behaviour during the “empirically-validated treatment” era [6.42]. It would not be expected that management – distant from the day to day clinical work yet held responsible for it - would prefer a treatment modality associated with
“boys on the roof” over a treatment modality associated with “boys not on the roof”. All other things being equal, the more difficult and costly option is unlikely to be supported without evidence of its greater effectiveness. As mentioned in the introduction, when judgements about mental health initiative are made in the absence of firm evidence of its effectiveness but in the presence of detailed information about its cost, lower-cost options will appear more attractive.

On the other hand, the long-term costs of ineffective intervention need also to be considered. The recent initiatives in this area have been made by economists. For example, the US economist James Heckman has argued that “American society underinvests in the very young and overinvests in mature adults with low skills” (2000, p.3). He bases his argument on studies of the later economic benefits of early emotional and academic intervention in infants, such benefits not always being apparent in the short term. While short term costs are obvious, the value of interventions can only be judged in the context of the later economic and other social costs of failing to adequately respond to youths who may indicate, in an unmistakable way through their severe problem behaviours both at home and at school, that their developmental needs are not being met.

The children aged ten at the commencement of the study in 1990 are now young adults twenty six years old. There is theoretical and clinical interest in examining the social outcomes for these young adults. What proportion remains in roles of angry defiance, or alienation, with their families or in institutions? What proportion is living independently, in satisfying relationships? What proportion has been incarcerated for violent acts against others? On the other hand, what proportion show outcomes which suggest, for example, that the child’s earlier angry defiance of their family was a healthy response to an unhealthy environment, a “protective” factor? That is, it may not necessarily be the case that those children showing the best short-term outcomes (as
investigated in the present study) will show the best long-term outcomes. Of particular interest will be the social and emotional adjustment of our former patients’ own children.

This is also relevant to the other important finding for the severe-pervasive problem subgroup. Better outcomes were achieved with children from step families than from other family structures, including two natural parents. As argued, this is important in developing an effective approach to severe child problem behaviours which may be associated with the transition to a “blended” or “reconstituted” family, transitions on the increase in the Western world. The impact of successful treatment outcomes for children of this subgroup, where the families have been helped to resolve problems associated with this transition, upon their later capacity as adults to manage similar problems is of particular interest. Given the more ambitious aims of psychodynamic milieu treatment, long-term outcomes may represent an opportunity to test them. The McCord (1978) study, discussed in detail [2.22], revisited youths thirty years later. In this way, while milieu therapy for severely emotionally disturbed children and adolescents might have ceased to be practiced, the evaluation of its effects can continue. The present short-term results justify further followup of these 159 lives in treatment. It may the case that the value of the evaluative research program from 1990 to 1999 is yet to fully emerge.

While the present results are of sufficient interest to justify further followup, some of the limitations of the original data collection (under the Arndell Evaluative Research Program) which emerged during the present study need to be considered. As mentioned, an ambiguity in the interpretation of the results with regard to step families arose from possible confounding with chronicity of the child’s problem. Information with regard to chronicity (i.e. duration of child problems preceding the introduction of the step-parent) may yet be collectable in a structured format from the families (e.g. the age at which their child first came to the attention of
the school counsellor) and checked against archival records (e.g. school files). As with the additions to the original data collection for the present study, taken from the clinical files (listed in Table 9), the limitations of the original data set need not restrict the scope of future research studies.

It is in this context that the striking and unexpected finding that was made with regard to outcomes across the entire study sample, and not just the severe-pervasive subgroup, need to be considered. The finding that children without severe problem behaviours at school were, on average, reported by their parents to keep improving between 6 months post-discharge and 12 months post-discharge, while children with severe problem behaviours at school were, on average, not reported to improve beyond the levels achieved at 6 months post-discharge, has been discussed in relation to the possibility that the latter group might have experienced additional problems, perhaps of an academic nature, which were not adequately addressed in the Unit in either treatment era. Information with regard to learning disabilities may be collectable in a structured format from the now adult ex-patients (e.g. formal tests of academic abilities and achievement). As discussed, collecting this information from these children prior to and during admission would have been very difficult, even if it had been a priority of the Unit at the time. It needs to be emphasised that this finding only arose due to the sub-grouping which was selected for the present data analysis, and that possible interactions between academic difficulties and behavioural difficulties have been largely neglected in the studies reviewed. For example, the academic deterioration in the more able students attending intensive treatment programs, discernible in the data presented in the papers, appears to have remained unrecognised [3.23, 3.41].
Other variables which may be extracted from the clinical files for further studies include the therapeutic alliance between the family and the Unit. As discussed [3.23], this is a treatment variable which is not only important to clinicians but has now accumulated a very substantial literature demonstrating its association with good outcomes across different treatment modalities. Grizenko (1997) demonstrated a very simple and elegant way to derive this information - the proportion of family sessions attended. To take another example, it has been discussed in relation to the differences in the philosophy of Unit treatment pre and post June 1996 [6.42] that there was a change over the course of the study in the attitude to medication. Structured information regarding the actual prescription of medication would be readily extractable from the clinical files. Yet another treatment variable not included in the present study was the length and quality of outpatient contact between the Unit and the family following the child’s discharge. The length of contact should also be readily extractable from the clinical files, and structured information with regard to the quality of outpatient contact may be collectable in a way similar to Grizenko’s method for assessing therapeutic alliance.

While the number of subjects limited the number of predictor variables included in the present data analyses, it needs to be emphasised that the present study extracted only a portion of the data from only a portion of the patients participating in the Arndell Evaluative Research Program, as outlined in the criteria for inclusion in the study [4.4]. The present study is not exhaustive of the study sample, much less the larger data set. A different research question involving different predictor or outcome variables would lead to the extraction of a different study sample. The larger numbers available from a less restricted extraction may allow, for example, a larger number of predictor variables to be included in the data analyses (although the questions which can be asked of that study sample will be different).
With regard to studies using an expanded or different set of predictor variables, it is acknowledged that any regression solution is very sensitive to the combination of variables included in it. That is, the selection of another set of predictors may produce quite different, equally legitimate, perspectives on the underlying structure of the data. Further possible studies of interest include a comparison of the study population against broader population data (e.g. census data) with regard to pre-treatment classificatory and predictor variables. The question asked may be, what are the predictors of admission to treatment? This is of particular interest with regard to examining possible selective processes operating within the Unit’s admission procedure [5.3]. To take a different example, an important feature of the data analysis in the present study is that the multiple linear regression analyses were conducted within clinically meaningful sub-groups. The selection of the present sub-grouping, based on the notion of taking differences between parent and teacher reports of problem behaviours as clinically meaningful (discussed in detail in 2.52), was validated by the unexpected finding with regard to the interaction between school-based behaviours and the trajectory of home-based outcomes. Yet, other clinically meaningful sub-groupings are possible (e.g. sub-grouping based on academic abilities). The models of predictors to emerge from analyses based on other sub-groupings may also produce quite different, equally legitimate, perspectives on the underlying structure of the data.

While these and other research projects building on the data collected under the Arndell Evaluative Research Program could keep researchers busy over a professional lifetime, the basic limitation of all such projects need to be recognised. It remains an N=1 study; it is a study of what took place over one time period in one treatment centre. Consistent with the advocacy of the naturalistic study of actual treatment settings and the pluralistic approach to research method throughout this thesis, it is argued that scientific aims will be best served not through
“replication” of the present study, whatever that might mean, but through independent, critically-minded evaluative research programs taking place at other settings. Randomised controlled trials of milieu treatment and cognitive-behavioural treatment, while welcome, are not anticipated.

From the perspective of policy making, it is of particular interest that a treatment variable emerged among the significant predictors of outcome in the most disturbed subgroup. This might be seen to offer some balance to the often nihilistic finding which emphasises the prominence of pre-treatment demographic and clinical factors, and the persistence of the effects of disadvantage, as discussed. For example, reference was made \textsuperscript{[2.51]} to the following statement from Kazdin and Weisz (1998) in the special section on “empirically supported psychological therapies” in the February 1998 issue of the \textit{Journal of Consulting and Clinical Psychology}. The statement was made with regard to predictors of outcome for problem-solving skills training for children showing externalizing behaviour.

“Some evidence suggests that... youths who have comorbid diagnoses, academic delays and dysfunction, and lower reading achievement and who come from families with high levels of impairment (parent psychopathology, stress, and family dysfunction) \textbf{respond less well to treatment} than youths with less dysfunction in these domains (Kazdin, 1995; Kazdin & Crowley, 1997)” (p. 25, emphasis added).

In fact, as was pointed out in Chapter 2, the study to which they refer (Kazdin & Crowley, 1997) made the unremarkable finding that those showing the most difficulties pre-treatment showed the most difficulties post-treatment. Kazdin & Crowley, however, proceeded to draw the following conclusions from their findings:
“In relation to clinical care the success of treatment will not only depend on identifying more effective interventions, but also on directing cases to available treatments from which they are likely to profit” (p. 205).

The discomforting implications of these types of statements for policies that may direct the most disturbed children - from the most disadvantaged family situations - away from treatment, has been discussed. Treatment may also be taken away from this group of children: the present results highlight the danger of poor research and poor research policy providing justification for closing down an effective treatment.

The present study, also using multiple linear regression analyses but in a way which highlighted (rather than de-emphasized) the therapeutic change taking place in the most disturbed subgroup, showed that there was a dynamic interaction between treatment, family and personal variables which accounted for a large amount of variation in outcomes in the severe-pervasive problem subgroup. A nihilistic view of the responsiveness of this subgroup to treatment is not supported. Rather, the present study suggested that this subgroup may be particularly responsive to psychodynamic milieu therapy, and that those from step-families may specially benefit. The policy implications of the present findings are quite different.

It remains to be asked, however, whether the present study fulfils the criteria for “empirical validation”. It is neither a random controlled trial nor a single-case experiment.

A distinct change appears to have taken place in this regard since the commencement of the project from which this study arises. Between the original submission of this thesis in September
2004 and the present revision, the American Psychological Association has released a policy statement on evidence-based practice in psychology (American Psychological Association, July 2005). Absent from this policy is the narrowly prescriptive tone with regard to research methodology and empirical validation found in the report of the Association’s Task Force on Promotion and Dissemination of Psychological Procedures (Task Force, 1995), a tone strongly evident in many of the contributions to the special edition of the Journal of Consulting and Clinical Psychology of February 1998, as discussed [2.3]. The 2005 policy statement contains a ready acknowledgement of “multiple types of research evidence”, of gaps and limitations in the existing literature, and an explicit recognition that “it is important not to assume that interventions that have not yet been studied in controlled trials are ineffective” (p. 7, op cit). It also states that “widely used psychological practices as well as innovations developed in the field or laboratory should be rigorously evaluated and barriers to conducting this research should be identified and addressed” (p. 7, op cit).

It appears then that the present study represents an evaluative research program in the field which is more consistent with the understanding of methodology now emerging than that which prevailed at the time the program commenced. It has identified and addressed barriers to conducting research, and fills a gap in the existing literature. It demonstrates a program which is modest in cost, yet sufficiently robust to have survived in the setting for which it was designed, producing substantive results with sufficient rigour to hold clinical and theoretical interest.
REFERENCES


Achenbach TM, Edelbrock C (1981) Behavioural problems and competencies reported by normal and disturbed children aged four through sixteen. *Monographs of the Society for Research in Child Development* No.188.


APPENDIX A

Approval of study from Royal North Shore Hospital
Medical Research Ethics Committee
Dear Mr Luiker,

Re: RNSH Protocol No. 9309-114(M) - H Luiker, N Kowalenko, B McDonald, G Sawyer

The Arndell Evaluative Research Programme.

I am pleased to inform you that the Medical Research Ethics Committee, at its meeting on 20 October 1993, approved your protocol on the above study.

In order to comply with the Guidelines for Good Clinical Research Practice (GCRP) in Australia, and in line with RNSH MREC policy, may I remind you that it is the Chief Investigator’s responsibility to ensure that:

i) A report is provided to the MREC at the completion of the study.

ii) The MREC is notified as soon as possible of any changes to the protocol. All changes must be approved by the MREC before continuation of the research project. This includes notifying the MREC of any changes to the staff involved with the protocol.

Yours sincerely,

REVEREND KEITH LITTLE
Acting Chairman
Medical Research Ethics Committee
APPENDIX B

Approval of study from Royal North Shore Hospital Medical Research Ethics Committee (RADGAC)
Mr H Luiker
Arndell Child & Adolescent Unit
Badajoz Road
North Ryde NSW 2113

Dear Mr Luiker,

Re: Protocol 9309-114M - H Luiker, N Kowalenko, B McDonald, G Sawyer
The Arndell evaluative research programme

Thank you for providing a copy of your RADGAC application for funding. The proposed project has previously been approved by the Medical Research Ethics Committee (MREC) under the title and protocol number above. I am able to extend this approval for a further three years.

In order to comply with the Guidelines for Good Clinical Research Practice (GCRP) in Australia, and in line with RNSH MREC policy, may I remind you that it is the chief investigator’s responsibility to ensure that:

1. A report is provided to the MREC at the completion of the study.
2. The MREC is notified as soon as possible of any changes to the protocol. All changes must be approved by the MREC before continuation of the research project. This includes notifying the MREC of any changes to the staff involved with the protocol.

Yours sincerely

GILLIAN M SHENFIELD
Chairman
Medical Research Ethics Committee
APPENDIX C

Child Behavior Checklist
**CHILD BEHAVIOR CHECKLIST FOR AGES 4-18**

**CHILD'S NAME**  Catherine

**SEX**  
- [ ] Boy  
- [x] Girl

**AGE**  14

**ETHNIC GROUP OR RACE**  

**TODAY'S DATE** 7/24/94

**CHILD'S BIRTHDATE**  
- Mo: 9  
- Date: 14  
- Year: 94

**GRADE IN SCHOOL**  9th

**NOT ATTEND AN S CHOOL**  

**PARENTS' USUAL TYPE OF WORK**, even if not working now. (Please be specific for example, auto mechanic, high school teacher, nurse, laborer, farm operator, shoe salesman, army sergeant)

**FATHER'S TYPE OF WORK**  Charter Officer

**MOTHER'S TYPE OF WORK**  Housewife

**THIS FORM FILLED OUT BY**  
- [ ] Mother  
- [ ] Father  
- [ ] Other - name & relationship to child

---

**I. Please list the sports your child most likes to take part in.**

- [ ] None
  - a. **Swimming**
  - b. **Tennis**
  - c. **Bike Riding**

---

**II. Please list your child's favorite hobbies, activities, and games, other than sports.**

For example: stamping, doll making, piano, crafts, cats, singing, etc. (Do not include listening to radio or TV)

- [ ] None
  - a. **Reading Books (especially mystery)**
  - b. **Music**
  - c. **Talking to Neighbors Daughter**

---

**III. Please list any organizations, clubs, teams, or groups your child belongs to.**

- [ ] None
  - a.  
  - b.  
  - c.  

---

**IV. Please list any jobs or chores your child has.**

For example: paper route, babysitting, making bed, working in store, etc. (Include both paid and unpaid jobs and chores)

- [ ] None
  - a. **Lunch Dishes Washing**
  - b. **Making bed and tidying room**
  - c. **Holiday House Work**

---

Copyright 1991 T.M. Achenbach, U. of Vermont,  
1 S. Prospect St, Burlington, VT 05401  
UNAUTHORIZED REPRODUCTION FORBIDDEN BY LAW  
1-91 Edition
V. 1. About how many close friends does your child have? □ None □ 1 □ 2 or 3 □ 4 or more
   (Do not include brothers & sisters)

2. About how many times a week does your child do things with any friends outside of regular school hours?
   (Do not include brothers & sisters) □ Less than 1 □ 1 or 2 □ 3 or more

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

<table>
<thead>
<tr>
<th>Worse</th>
<th>About Average</th>
<th>Better</th>
<th>□ Has no brothers or sisters</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Get along with his/her brothers &amp; sisters?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Get along with other kids?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Behave with his/her parents?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Play and work by himself/herself?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

VII. 1. For ages 6 and older—performance in academic subjects. If child is not being taught, please give reason

<table>
<thead>
<tr>
<th>Failing</th>
<th>Below average</th>
<th>Average</th>
<th>Above average</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Reading, English, or Language Arts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. History or Social Studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Arithmetic or Math</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Science</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Religious Studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. Physical Education &amp; Sports</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g. Geography</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Other academic subjects—for example: computer courses, foreign language, business. Do not include gym, shop, driver's ed, etc.

2. Is your child in a special class or special school? □ No □ Yes—what kind of class or school?

3. Has your child 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?

Does your child have any illness, physical disability, or mental handicap? □ No □ Yes—please describe

What concerns you most about your child?

Mood changes very easily. Anger can be triggered by something minor and she becomes very aggressive physically towards me and our son (10 years old). She is very jealous of our son. Her aggressive behavior and reluctance to go to school is of great concern.

Please describe the best things about your child:

She always apologises and says sorry for her outburst of temper. This usually occurs hours after the event. She can be affectionate when she wants to.
Below is a list of items that describe children and youth. 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

<table>
<thead>
<tr>
<th>Item</th>
<th>Code</th>
<th>Very True or Often True</th>
<th>Somewhat or Sometimes True</th>
<th>Not True (as far as you know)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0 1 2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Acts too young for his/her age</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Allergy (describe):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argues a lot</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Asthma</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Behaves like opposite sex</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Bowel movements outside toilet</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Bragging, boasting</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Can’t concentrate, can’t pay attention for long</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Can’t get his/her mind off certain thoughts; obsessions (describe):</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Can’t sit still, restless, or hyperactive</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Clings to adults or too dependent</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Complains of loneliness</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Confused or seems to be in a fog</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Cries a lot</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Cruel to animals</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Cruelty, bullying, or meannessanness to others</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Day-dreams or gets lost in his/her thoughts</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Deliberately harms self or attempts suicide</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Demands a lot of attention</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Destroys his/her own things</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Destroys things belonging to his/her family or others</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Disobedient at home</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Disobedient at school</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Doesn’t eat well</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Doesn’t get along with other kids</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Doesn’t seem to feel guilty after misbehaving</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Easily jealous</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Eats or drinks things that are not food — don’t include sweets (describe):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fears certain animals, situations, or places, other than school (describe):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fears going to school</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Feels he/she might think or do something bad</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Feels he/she has to be perfect</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Feels or complains that no one loves him/her</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Feels others are out to get him/her</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Feels worthless or inferior</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Gets hurt a lot, accident-prone</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Gets in many fights</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Gets teased a lot</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Hangs around with others who get in trouble</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Hears sounds or voices that aren’t there (describe): <strong>when alone at home</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impulsive or acts without thinking</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Lying or cheating</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Bites fingernails</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Nervous, high-strung, or tense</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Nervous movements or twitching (describe): <strong>twitching fingers especially when tired</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nightmares</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Not liked by other kids</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Constipated, doesn’t move bowels</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Too fearful or anxious</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Feels dizzy</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Feels too guilty</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Overeating, especially junk food like ice cream, sweets</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Over tired</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Overweight</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Physical problems without known medical cause:</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>a. Aches or pains (not headaches)</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>b. Headaches not so much now</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>c. Nausea, feels sick</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>d. Problems with eyes (describe):</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>e. Rashes or other skin problems</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>f. Stomachaches or cramps</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>g. Vomiting, throwing up</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>h. Other (describe):</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0 = Not True (as far as you know)</td>
<td>1 = Somewhat or Sometimes True</td>
<td>2 = Very True or Often True</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------</td>
<td>--------------------------------</td>
<td>--------------------------</td>
<td></td>
</tr>
<tr>
<td>57.</td>
<td>Physically attacks people verbally or physically.</td>
<td></td>
<td>0 1 2 84. Strange behavior (describe): She talks to herself.</td>
<td></td>
</tr>
<tr>
<td>58.</td>
<td>Picks nose, skin, or other parts of body. (describe): She used to pick her nose. Scabs are found on nails but she denies it now.</td>
<td></td>
<td>0 1 2 85. Strange ideas (describe): Feelings of insecurity, not wanted or loved. How would you ever tell me to leave?</td>
<td></td>
</tr>
<tr>
<td>59.</td>
<td>Plays with own sex parts in public.</td>
<td></td>
<td>0 1 2 86. Stubborn, sullen, or irritable.</td>
<td></td>
</tr>
<tr>
<td>60.</td>
<td>Plays with own sex parts too much.</td>
<td></td>
<td>0 1 2 87. Sudden changes in mood or feelings.</td>
<td></td>
</tr>
<tr>
<td>61.</td>
<td>Poor school work.</td>
<td></td>
<td>0 1 2 88. Sukoa a lot.</td>
<td></td>
</tr>
<tr>
<td>62.</td>
<td>Poorly coordinated or clumsy.</td>
<td></td>
<td>0 1 2 89. Suspicious.</td>
<td></td>
</tr>
<tr>
<td>63.</td>
<td>Prefers being with older kids.</td>
<td></td>
<td>0 1 2 90. Swearing or obscene language.</td>
<td></td>
</tr>
<tr>
<td>64.</td>
<td>Prefers being with younger kids.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65.</td>
<td>Refuses to talk.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66.</td>
<td>Repeats certain acts over and over. Compulsions (describe): Always costs five dollars for the ticket.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>67.</td>
<td>Runs away from home.</td>
<td></td>
<td>0 1 2 91. Talks about killing self.</td>
<td></td>
</tr>
<tr>
<td>68.</td>
<td>Screams a lot.</td>
<td></td>
<td>0 1 2 92. Talks or walks in sleep (describe): Talks aloud in dreams.</td>
<td></td>
</tr>
<tr>
<td>69.</td>
<td>Secretive, keeps things to self.</td>
<td></td>
<td>0 1 2 93. Talks too much.</td>
<td></td>
</tr>
<tr>
<td>70.</td>
<td>Sees things that aren’t there (describe):</td>
<td></td>
<td>0 1 2 94. Teases a lot.</td>
<td></td>
</tr>
<tr>
<td>71.</td>
<td>Self-conscious or easily embarrassed.</td>
<td></td>
<td>0 1 2 95. Temper tantrums or hot temper.</td>
<td></td>
</tr>
<tr>
<td>72.</td>
<td>Sets fires. (describe): Setting fire to the furniture.</td>
<td></td>
<td>0 1 2 96. Thinks about sex too much.</td>
<td></td>
</tr>
<tr>
<td>73.</td>
<td>Sexual problems (describe):</td>
<td></td>
<td>0 1 2 97. Threatens people.</td>
<td></td>
</tr>
<tr>
<td>74.</td>
<td>Showing off or clowning.</td>
<td></td>
<td>0 1 2 98. Thumb-sucking.</td>
<td></td>
</tr>
<tr>
<td>75.</td>
<td>Shy or timid.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76.</td>
<td>Sleeps less than most kids.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77.</td>
<td>Sleeps more than most kids during day and/or night (describe): She sleeps in the morning after being up late the night before.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>78.</td>
<td>Smears or plays with bowel movements.</td>
<td></td>
<td>0 1 2 99. Too concerned with neatness or cleanliness.</td>
<td></td>
</tr>
<tr>
<td>79.</td>
<td>Speech problem (describe): Sometimes her thoughts are not completed before she speaks.</td>
<td></td>
<td>0 1 2 100. Trouble sleeping (describe): Before sleeping she has to take a glass of water and drink it before bedtime.</td>
<td></td>
</tr>
<tr>
<td>80.</td>
<td>Stares blankly.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>81.</td>
<td>Steals at home.</td>
<td></td>
<td>0 1 2 101. Truancy, skips school.</td>
<td></td>
</tr>
<tr>
<td>82.</td>
<td>Steals outside the home.</td>
<td></td>
<td>0 1 2 102. Underactive, slow moving, or lacks energy.</td>
<td></td>
</tr>
<tr>
<td>83.</td>
<td>Stores up things he/she doesn’t need (describe):</td>
<td></td>
<td>0 1 2 103. Unhappy, sad, or depressed.</td>
<td></td>
</tr>
<tr>
<td>0 1 2 104. Unusually loud.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>105.</td>
<td>Uses alcohol or drugs for nonmedical purposes (describe):</td>
<td></td>
<td>0 1 2 106. Vandalism.</td>
<td></td>
</tr>
<tr>
<td>107.</td>
<td>Wets self during the day.</td>
<td></td>
<td>0 1 2 108. Wets the bed.</td>
<td></td>
</tr>
<tr>
<td>109.</td>
<td>Whining.</td>
<td></td>
<td>0 1 2 109. Wishes to be of opposite sex.</td>
<td></td>
</tr>
<tr>
<td>110.</td>
<td>Wishes to be of opposite sex.</td>
<td></td>
<td>0 1 2 111. Withdrawn, doesn’t get involved with others.</td>
<td></td>
</tr>
<tr>
<td>112.</td>
<td>Worries.</td>
<td></td>
<td>0 1 2 112. Worrying.</td>
<td></td>
</tr>
</tbody>
</table>

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

0 1 2

Please be sure you have answered all items.

Underline any you are concerned about.
APPENDIX D

Teacher Report Form
TEACHER'S REPORT FORM

Your answers will be used to compare the pupil with other pupils whose teachers have completed similar forms. The information from this form will also be used for comparison with other information about this pupil. Please answer as well as you can, even if you lack full information. Scores on individual items will be combined to identify general patterns of behavior. Feel free to write additional comments beside each item and in the spaces provided on page 2.

**PUPIL'S NAME**

MATTHEW

**PUPIL'S SEX**

☑ Boy ☐ Girl

**PUPIL'S AGE**

9

**ETHNIC GROUP**

AUSTRALIAN

**OR RACE**

**PARENTS' USUAL TYPE OF WORK, even if not working now. (Please be as specific as you can—e.g., auto mechanic, high school teacher, housewife, laborer, first aid operator, shoe salesman, army sergeant.)**

**FATHERS' TYPE OF WORK:** UNKNOWN

**MOTHERS' TYPE OF WORK:** NURSE

**TODAY'S DATE**

May 30, 1994

**PUPIL'S BIRTHDATE (if known)**

**GRADE IN SCHOOL**

4

**NAME OF SCHOOL**

YATES AVE

**SD**

1

**SD**

I. How long have you known this pupil?

12 months

II. How well do you know this pupil?

1. ☐ Not Well

2. ☑ Moderately Well

3. ☐ Very Well

III. How much time does he/she spend in your class per week?

1993 - Normal lesson time.

IV. What kind of class is it? (Please be specific, e.g., regular 5th grade, 7th grade math, etc.)

1993 - Parallel.

V. Has he/she ever been referred for special class placement, services, or tutoring? ☐ Don't Know ☐ Yes — what kind and when?

☑ Part-time integration

☐ Full-time integration

☑ Redsh Email placement sought. Term 4.

VI. Has he/she ever repeated a grade?

☑ Don't Know ☐ No

☑ Yes — grade and reason

VII. Current school performance — list academic subjects and check column that indicates pupil's performance:

**Nov. 1993**

<table>
<thead>
<tr>
<th>Academic subject</th>
<th>1. Far below grade</th>
<th>2. Somewhat below grade</th>
<th>3. At grade level</th>
<th>4. Somewhat above grade</th>
<th>5. Far above grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ENGLISH</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>2. MATHS</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>3. SCIENCE &amp; TECHNOLOGY</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>4. HUMAN SOCIETY &amp; ITS</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>5. CREATIVE &amp; PRACTICAL</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>6. HEALTH/P.E./P.D.</td>
<td>☐</td>
<td>☐</td>
<td>☑</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

*Copyright 1991 Thomas M. Achenbach
Center for Children, Youth & Families
University of Vermont
1 South Prospect St.
Burlington, VT 05401

UNAUTHORIZED REPRODUCTION FORBIDDEN BY LAW

1/91 Edition
VIII. Compared to typical pupils of the same age:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>How hard is he/she working?</td>
<td>□</td>
<td>✔</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>How appropriately is he/she behaving?</td>
<td>✔</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>How much is he/she learning?</td>
<td>□</td>
<td>✔</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>How happy is he/she?</td>
<td>□</td>
<td>✔</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

IX. Most recent achievement test scores (if available):

<table>
<thead>
<tr>
<th>Name of test</th>
<th>Subject</th>
<th>Date</th>
<th>Percentile or grade level obtained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading Grade Test</td>
<td>Reading</td>
<td>Nov. 1993</td>
<td>C grading</td>
</tr>
<tr>
<td>Number Grade Test</td>
<td>Maths</td>
<td>Nov. 1993</td>
<td>83% C grading</td>
</tr>
<tr>
<td>Measurement</td>
<td>Maths</td>
<td>Nov. 1993</td>
<td>78% C grade level</td>
</tr>
<tr>
<td>Handwriting</td>
<td>Writing</td>
<td>Nov. 1993</td>
<td>C grading</td>
</tr>
</tbody>
</table>

X. IQ, readiness, or aptitude tests (if available):

<table>
<thead>
<tr>
<th>Name of test</th>
<th>Date</th>
<th>IQ or equivalent scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Skills Test 1993</td>
<td>August 1993</td>
<td>150</td>
</tr>
<tr>
<td>Literacy skill level</td>
<td>49</td>
<td>Numeracy skill level</td>
</tr>
</tbody>
</table>

Does this pupil have any illness, physical disability, or mental handicap? □ No □ Yes – please describe

What concerns you most about this pupil?

Disruptive and attention seeking behaviour - non-compliance. Attention seeking behaviour manifested itself in the following ways: incessant sounds; answering back; swearing; deliberate hurting others; throwing objects; kicking - physical aggression to students and teachers.

Please describe the best things about this pupil:

Class teacher found this difficult. I support her on this. Matthew was one of the very rare numbers about whom one could not really find anything particularly likeable. As a professional this does not rest well. F.Y.F.W. (Proverb)

Please feel free to write any comments about this pupil’s work, behavior, or potential, using extra pages if necessary.

From results of Basic Skills and personal observations, I feel Matthew's academic progress was adversely affected by his poor social behaviour.
Below is a list of items that describe pupils. For each item that describes the pupil now or within the past 2 months, please circle the 2 if the item is very true or often true of the pupil. Circle the 1 if the item is somewhat or sometimes true of the pupil. If the item is not true of the pupil, circle the 0. Please answer all items as well as you can, even if some do not seem to apply to this pupil.

<table>
<thead>
<tr>
<th>Item</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.2</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.4</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.6</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.7</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.8</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.9</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1.1</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1.2</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1.3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1.4</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1.5</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1.6</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1.7</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1.8</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1.9</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.1</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.2</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.4</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.5</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.6</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.7</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.8</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.9</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

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

---

**Additional Items:**
- Eats or drinks things that are not food (describe)
- Fears certain animals, situations, or places other than school (describe)
- Fears going to school

---

**Please see other side**
<table>
<thead>
<tr>
<th>0 = Not True (as far as you know)</th>
<th>1 = Somewhat or Sometimes True</th>
<th>2 = Very True or Often True</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 57. Physically attacks people</td>
<td>1 2 84. Strange behavior (describe):</td>
<td></td>
</tr>
<tr>
<td>0 1 2 56. Picks nose, skin, or other parts of body (describe):</td>
<td>1 2 85. Strange ideas (describe):</td>
<td></td>
</tr>
<tr>
<td>0 1 2 59. Sleeps in class</td>
<td>0 1 2 86. Stubborn, sullen, or irritable</td>
<td></td>
</tr>
<tr>
<td>0 1 2 60. Apathetic or unmotivated; was quite competitive at times in morning spelling.</td>
<td>0 1 2 87. Sudden changes in mood or feelings</td>
<td></td>
</tr>
<tr>
<td>0 1 2 61. Poor school work</td>
<td>0 1 2 88. Sucks a lot</td>
<td></td>
</tr>
<tr>
<td>0 1 2 62. Poorly coordinated or clumsy; poor coordination with sports activities</td>
<td>0 1 2 89. Suspicious</td>
<td></td>
</tr>
<tr>
<td>0 1 2 63. Prefers being with older children or youths</td>
<td>0 1 2 90. Swearing or obscene language</td>
<td></td>
</tr>
<tr>
<td>0 1 2 64. Prefers being with younger children</td>
<td>0 1 2 91. Talks about killing self</td>
<td></td>
</tr>
<tr>
<td>0 1 2 65. Refuses to talk</td>
<td>0 1 2 92. Underachieving, not working up to potential</td>
<td></td>
</tr>
<tr>
<td>0 1 2 66. Repeats certain acts over and over; compulsions (describe); incessant noises/sounds in class; burning, itching, rolling on the ground for attention</td>
<td>0 1 2 93. Talks too much</td>
<td></td>
</tr>
<tr>
<td>0 1 2 67. Disrupts class discipline</td>
<td>0 1 2 94. Teases a lot</td>
<td></td>
</tr>
<tr>
<td>0 1 2 68. Screams a lot</td>
<td>0 1 2 95. Temper tantrums or hot temper</td>
<td></td>
</tr>
<tr>
<td>0 1 2 69. Socratic, keeps things to self</td>
<td>0 1 2 96. Seems preoccupied with sex</td>
<td></td>
</tr>
<tr>
<td>0 1 2 70. Sees things that aren't there (describe):</td>
<td>0 1 2 97. Threatens people; also feels like</td>
<td></td>
</tr>
<tr>
<td>0 1 2 71. Self-conscious or easily embarrassed</td>
<td>0 1 2 98. Tardy to school or class; the &quot;victim&quot;</td>
<td></td>
</tr>
<tr>
<td>0 1 2 72. Messy work; poor handwriting, setting out etc.</td>
<td>0 1 2 99. Too concerned with neatness or cleanliness</td>
<td></td>
</tr>
<tr>
<td>0 1 2 73. Behaves irresponsibly (describe):</td>
<td>0 1 2 100. Falls to carry out assigned tasks</td>
<td></td>
</tr>
<tr>
<td>0 1 2 74. Showing off or clowning</td>
<td>0 1 2 101. Truanty or unexplained absence</td>
<td></td>
</tr>
<tr>
<td>0 1 2 75. Shy or timid</td>
<td>0 1 2 102. Underactive, slow moving, or lacks energy</td>
<td></td>
</tr>
<tr>
<td>0 1 2 76. Explosive and unpredictable behavior; erratic</td>
<td>0 1 2 103. Unhappy, sad, or depressed</td>
<td></td>
</tr>
<tr>
<td>0 1 2 77. Demands must be met immediately, easily frustrated</td>
<td>0 1 2 104. Unusually loud</td>
<td></td>
</tr>
<tr>
<td>0 1 2 78. Inattentive, easily distracted; depends on the teacher</td>
<td>0 1 2 105. Uses alcohol or drugs for nonmedical purposes (describe):</td>
<td></td>
</tr>
<tr>
<td>0 1 2 79. Speech problem (describe):</td>
<td>0 1 2 106. Overly anxious to please</td>
<td></td>
</tr>
<tr>
<td>0 1 2 80. Blares blankly</td>
<td>0 1 2 107. Dislikes school</td>
<td></td>
</tr>
<tr>
<td>0 1 2 81. Feels hurt when criticized; difficult to ascertain</td>
<td>0 1 2 108. Is afraid of making mistakes</td>
<td></td>
</tr>
<tr>
<td>0 1 2 82. Steals</td>
<td>0 1 2 109. Whining</td>
<td></td>
</tr>
<tr>
<td>0 1 2 83. Stores up things he/she doesn't need (describe):</td>
<td>0 1 2 110. Unclean personal appearance</td>
<td></td>
</tr>
<tr>
<td>0 1 2 84. Strangers behavior (describe):</td>
<td>0 1 2 111. Withdrawn; doesn't get involved with others</td>
<td></td>
</tr>
<tr>
<td>0 1 2 85. Strange ideas (describe):</td>
<td>0 1 2 112. Worries</td>
<td></td>
</tr>
<tr>
<td>0 1 2 86. Stubborn, sullen, or irritable</td>
<td>0 1 2 113. Please write in any problems the pupil has that were not listed above:</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX E

*Family Assessment Device*
Version 3
Nathan B. Epstein, M.D.
Lawrence M. Baldwin, Ph.D.
Duane S. Bishop, M.D.
The Brown University/Butler Hospital Family Research Program
Butler Hospital
345 Blackstone Boulevard
Providence, Rhode Island 02906

Date of Administration: ____________________________

Family Role: ______________________________________

Identification Number or Family Name: ____________________________
INSTRUCTIONS:

This booklet contains a number of statements about families. Please read each statement carefully, and decide how well it describes your own family. You should answer according to how you see your family.

For each statement there are four (4) possible responses:

Strongly Agree (SA) Check SA if you feel that the statement describes your family very accurately.

Agree (A) Check A if you feel that the statement describes your family for the most part.

Disagree (D) Check D if you feel that the statement does not describe your family for the most part.

Strongly Disagree (SD) Check SD if you feel that the statement does not describe your family at all.

These four responses will appear below each statement like this:

41. We are not satisfied with anything short of perfection.

____ SA ____ A ____ D ____ SD

The answer spaces for statement 41 would look like this. For each statement in the booklet, there is an answer space below. Do not pay attention to the blanks at the far right-hand side of each answer space. They are for office use only.

Try not to spend too much time thinking about each statement, but respond as quickly and as honestly as you can. If you have trouble with one, answer with your first reaction. Please be sure to answer every statement and mark all your answers in the space provided below each statement.
1. Planning family activities is difficult because we misunderstand each other.
   ____ SA  ____ A  ____ D  ____ SD

2. We resolve most everyday problems around the house.
   ____ SA  ____ A  ____ D  ____ SD

3. When someone is upset the others know why.
   ____ SA  ____ A  ____ D  ____ SD

4. When you ask someone to do something, you have to check that they did it.
   ____ SA  ____ A  ____ D  ____ SD

5. If someone is in trouble, the others become too involved.
   ____ SA  ____ A  ____ D  ____ SD

6. In times of crisis we can turn to each other for support.
   ____ SA  ____ A  ____ D  ____ SD

7. We don't know what to do when an emergency comes up.
   ____ SA  ____ A  ____ D  ____ SD

8. We sometimes run out of things that we need.
   ____ SA  ____ A  ____ D  ____ SD

9. We are reluctant to show our affection for each other.
   ____ SA  ____ A  ____ D  ____ SD

10. We make sure members meet their family responsibilities.
    ____ SA  ____ A  ____ D  ____ SD

11. We cannot talk to each other about the sadness we feel.
    ____ SA  ____ A  ____ D  ____ SD

12. We usually act on our decisions regarding problems.
    ____ SA  ____ A  ____ D  ____ SD
13. You only get the interest of others when something is important to them.
   ____ SA ____ A ____ D ____ SD

14. You can't tell how a person is feeling from what they are saying.
   ____ SA ____ A ____ D ____ SD

15. Family tasks don't get spread around enough.
   ____ SA ____ A ____ D ____ SD

16. Individuals are accepted for what they are.
   ____ SA ____ A ____ D ____ SD

17. You can easily get away with breaking the rules.
   ____ SA ____ A ____ D ____ SD

18. People come right out and say things instead of hinting at them.
   ____ SA ____ A ____ D ____ SD

19. Some of us just don't respond emotionally.
   ____ SA ____ A ____ D ____ SD

20. We know what to do in an emergency.
   ____ SA ____ A ____ D ____ SD

21. We avoid discussing our fears and concerns.
   ____ SA ____ A ____ D ____ SD

22. It is difficult to talk to each other about tender feelings.
   ____ SA ____ A ____ D ____ SD

23. We have trouble meeting our bills.
   ____ SA ____ A ____ D ____ SD

24. After our family tries to solve a problem, we usually discuss whether it worked or not.
   ____ SA ____ A ____ D ____ SD
25. We are too self-centered.
   ____ SA _____ A _____ D _____ SD

26. We can express feelings to each other.
   ____ SA _____ A _____ D _____ SD

27. We have no clear expectations about toilet habits.
   ____ SA _____ A _____ D _____ SD

28. We do not show our love for each other.
   ____ SA _____ A _____ D _____ SD

29. We talk to people directly rather than through go-betweens.
   ____ SA _____ A _____ D _____ SD

30. Each of us has particular duties and responsibilities.
   ____ SA _____ A _____ D _____ SD

31. There are lots of bad feelings in the family.
   ____ SA _____ A _____ D _____ SD

32. We have rules about hitting people.
   ____ SA _____ A _____ D _____ SD

33. We get involved with each other only when something interests us.
   ____ SA _____ A _____ D _____ SD

34. There's little time to explore personal interests.
   ____ SA _____ A _____ D _____ SD

35. We often don't say what we mean.
   ____ SA _____ A _____ D _____ SD

36. We feel accepted for what we are.
   ____ SA _____ A _____ D _____ SD
37. We show interest in each other when we can get something out of it personally.
   ____ SA _____ A _____ D _____ SD

38. We resolve most emotional upsets that come up.
   ____ SA _____ A _____ D _____ SD

39. Tenderness takes second place to other things in our family.
   ____ SA _____ A _____ D _____ SD

40. We discuss who is to do household jobs.
   ____ SA _____ A _____ D _____ SD

41. Making decisions is a problem for our family.
   ____ SA _____ A _____ D _____ SD

42. Our family shows interest in each other only when they can get something out of it.
   ____ SA _____ A _____ D _____ SD

43. We are frank with each other.
   ____ SA _____ A _____ D _____ SD

44. We don't hold to any rules or standards.
   ____ SA _____ A _____ D _____ SD

45. If people are asked to do something, they need reminding.
   ____ SA _____ A _____ D _____ SD

46. We are able to make decisions about how to solve problems.
   ____ SA _____ A _____ D _____ SD

47. If the rules are broken, we don't know what to expect.
   ____ SA _____ A _____ D _____ SD

48. Anything goes in our family.
   ____ SA _____ A _____ D _____ SD
49. We express tenderness.
   _____ SA _____ A _____ D _____ SD

50. We confront problems involving feelings.
   _____ SA _____ A _____ D _____ SD

51. We don't get along well together.
   _____ SA _____ A _____ D _____ SD

52. We don't talk to each other when we are angry.
   _____ SA _____ A _____ D _____ SD

53. We are generally dissatisfied with the family duties assigned to us.
   _____ SA _____ A _____ D _____ SD

54. Even though we mean well, we intrude too much into each other's lives.
   _____ SA _____ A _____ D _____ SD

55. There are rules about dangerous situations.
   _____ SA _____ A _____ D _____ SD

56. We confide in each other.
   _____ SA _____ A _____ D _____ SD

57. We cry openly.
   _____ SA _____ A _____ D _____ SD

58. We don't have reasonable transport.
   _____ SA _____ A _____ D _____ SD

59. When we don't like what someone has done, we tell them.
   _____ SA _____ A _____ D _____ SD

60. We try to think of different ways to solve problems.
   _____ SA _____ A _____ D _____ SD
Instructions on How to Score the FAD

The McMaster Family Assessment Device (FAD) is designed to measure family functioning as described in the McMaster Model of Family Functioning. It is made up of seven scales, one measuring overall family functioning and one for each of the six dimensions of the McMaster Model. Each of the items on the FAD belongs to only one of the seven scales. Some items describe healthy functioning while others describe unhealthy functioning. Table 1 (below) indicates the items for each scale, classified according to whether they describe healthy or unhealthy functioning.

<table>
<thead>
<tr>
<th>Problem Solving</th>
<th>Communication</th>
<th>Roles</th>
<th>Affective Responsiveness</th>
<th>Affective Involvement</th>
<th>Behavior Control</th>
<th>General Functioning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>2*</td>
<td>3</td>
<td>10</td>
<td>49</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>Functioning</td>
<td>12</td>
<td>18</td>
<td>30*</td>
<td>57</td>
<td>32</td>
<td>12</td>
</tr>
<tr>
<td>Items</td>
<td>24</td>
<td>29*</td>
<td>40</td>
<td>55</td>
<td>55</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>43</td>
<td></td>
<td></td>
<td>36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>59</td>
<td></td>
<td></td>
<td>46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Unhealthy</td>
<td>14</td>
<td>4</td>
<td>9</td>
<td>5</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Functioning</td>
<td>22*</td>
<td>8*</td>
<td>19</td>
<td>13</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>Items</td>
<td>35*</td>
<td>15</td>
<td>25</td>
<td>27</td>
<td>27</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>39</td>
<td>33</td>
<td>44</td>
<td>44</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>42</td>
<td>48</td>
<td></td>
<td>48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>53</td>
<td>54</td>
<td></td>
<td></td>
<td>54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>58*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is important to note the items that are marked in Table 1 with an asterisk. These are the seven items added after the original report which increase the reliability of the three scales while not affecting their intercorrelation with other scales. We have used the 60-item version in all subsequent research and recommend others use this version.

To score the FAD, all responses are coded as follows:

- Strongly Agree = 1
- Agree = 2
- Disagree = 3
- Strongly Disagree = 4
Instructions on How to Score the FAD

Brown/Butler Family Research Program

Then the score for items describing unhealthy functioning are transformed by subtracting them from 5. This inverts the response scales on the unhealthy items and has the effect of equating a Strongly Agree response to an unhealthy item with a Strongly Disagree response to a healthy item, etc. As a result of this transformation, 1 represents a healthy response and 4 represents an unhealthy response for all items. These scored responses to the items of each scale are averaged to provide seven scale scores each having a possible range from 1.00 (healthy) to 4.00 (unhealthy).*

The manual FAD scoring sheet we have developed is included. It can be used to score an individual's responses if a computer scoring system is not used. The first step is to score all the answers in the column on the extreme left. The negative items (with an asterisk) are then transformed by subtracting them from 5 and entering them in the second column headed 'transformed score'.

On the right-hand side of the scoring sheet are seven columns of boxes, one column for each of the seven scales. The scale to which an item belongs is indicated by the column in which the box aligned with the item falls. The item scores (transformed scores for unhealthy items) are next transferred to their appropriate boxes. To calculate a scale score, simply add the scores in each column and divide the sum by the number of items in the column that were answered. Scale scores will range from 1.00 (healthy) to 4.00 (unhealthy).

We have also developed a computer program for scoring the FAD. This program provides individual scores and a family mean score, as well as item analyses. It can also be used to administer the FAD to subjects interactively. This scoring program is available for the Macintosh microcomputer, and is designed to be used with Microsoft Basic, which you must own.

If you prefer, the Brown/Butler Family Research Program has computer programs to score the FAD for you. These programs generate the seven dimension scale scores for individuals and some family measure scores. The family measure scores are experimental and we have not explored them in detail as of yet. They include median scores for the father, mother, patient, parent, non-parent, and non-patient family members along with maximum, median, and minimum scores for the family. We will score a family and provide a printout of the individual and family scores for a cost of $5.00 per family. We ask that you send the Family Information Form with the FAD data if you wish to have the data scored by us.

*If more than 40% of the items for a scale are missing, a scale score is not calculated, it is designated missing.
<table>
<thead>
<tr>
<th>Response</th>
<th>Transformed Score</th>
<th>Problem Solving</th>
<th>Communication</th>
<th>Roles</th>
<th>Affective Responsiveness</th>
<th>Affective Involvement</th>
<th>Behavior</th>
<th>General Control</th>
<th>Functioning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>5 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Unhealthy items which are subtracted from 5.
APPENDIX F

Consent form at registration
ARNDELL CHILD AND ADOLESCENT SERVICE
CONSENT FORM TO PARTICIPATE IN A RESEARCH PROJECT

I, ................................................ being the..........................................................
(State Relationship e.g. parent, guardian)
of ..........................................................
(name of patient)
of ..........................................................
(Address)
have been asked to consent to my relative/child's participation in a research project. I have been informed
of the following points:

1. The aim of this project is to evaluate the effectiveness of treatment at Arndell Child and
   Adolescent Unit.

2. The results which will be obtained may or may not be of direct benefit to the management of
   my relative/child.

3. The procedure will involve completion of questionnaires which ask about my child's
   behaviour and about how we are managing as a family. Requests to complete questionnaires
   will occur throughout the course of treatment and at 6 and 12 months following discharge.

4. I give my permission to access my child's medical and school records, and for
   the researchers to ask teachers about my child's behaviour and progress at school.

5. I understand that I or my relative/child can refuse to take part in this study or withdraw my
   permission at any time, and that this will not affect the treatment offered to me by the
   Department of Child and Adolescent Psychiatry, Royal North Shore Hospital.

6. Participation in this study will be at no financial cost to me.

7. My answers to the questionnaires and any information gained from records will remain
   completely confidential and no material may be published that can identify us.

8. Should I have any concerns or questions relating to my involvement in this project I am aware
   that I may contact Mr. Henry Luiker (02) 887.5830

9. After considering all of these points I consent to my, ................................................
    (state relationship)
    daughter/s
    participation in this study.

I also state that my relative/child has/have not participated in any other research project in
the last 3 months.

SIGNATURE............................................................

WITNESS............................................................
(Please print name)

DATE............................................................
9/5/3

SIGNATURE............................................................
(of witness)
APPENDIX G

Consent form at discharge
ARNDELL CHILD & ADOLESCENT UNIT

PARENTAL CONSENT

I, .................................................................................................................. being the ........................................................................................................ (State relationship to child)
of ........................................................................................................ (name of child)
of ........................................................................................................ (address) ........................................................................................................ (postcode)

have been invited to participate, and to allow my child to participate, in the Arndell Follow-up Study.

1: The aim of the project is to look at the progress of children/adolescents and their families after leaving Arndell Child & Adolescent Unit.

2: One part of the procedure will involve my family being asked to complete questionnaires, which will be posted to us, about our progress 6 months and 12 months after leaving Arndell.

3: Another part of the procedure will involve Arndell contacting my child's school to inquire about his/her progress 6 months and 12 months after leaving Arndell.

4: I understand that I can refuse to take part in this study, or refuse to have my child take part, or withdraw my permission at any time, and that this will not affect the treatment offered to me or my family at Arndell Child & Adolescent Unit.

5: Any information gained will remain completely confidential.

6: Should I have any concerns or questions relating to my involvement in this project, I am aware that I may contact Mr Henry Luiker at the Arndell Child & Adolescent Unit on (02)887–5830.

After considering all these points:

(A) I DO/DO NOT consent to having questionnaires sent to me.

(B) I DO/DO NOT consent to questionnaires being sent to my child's school.

(please circle your answers)

SIGNATURE .......................................................... ................................

DATE .......................................................... ...........................
APPENDIX H

Calculation of Jacobson’s Reliable Change Index (RCI) for the Child Behavior Checklist Total Problem Score.

As outlined [2.41], the standard error of measurement can be computed from the standard deviation of test scores in the normal population and the test-retest reliability.

\[ S_E = sd\sqrt{1 - r_{xx}} \]

From the contemporaneous manual (CBCL/4-18 Profile for Boys – Problem Scales, 1991, T.M.Achenbach.) the Total Problem Score has a test-retest reliability of .93 at one-week (Table 5-1, p. 73), and the standard deviation of 16.4 in the normal population can be calculated by averaging across the four age/gender subgroups appearing in Appendix B of the manual.

\[ S_E = 16.4\sqrt{1 - .93} \]

This gives a standard error for the Total Problem Score of 4.339.

Standard error of the difference between the pre-treatment and post-treatment test scores is computed from the standard error of measurement.

\[ S_{\text{diff}} = \sqrt{2(S_E)^2} \]
\[ S_{\text{diff}} = \sqrt{2(4.339)^2} \]
\[ S_{\text{diff}} = 6.13 \]

The RCI is defined as the difference between an individual’s pretest score and posttest score divided by the standard error of the difference between the two test scores.
Where the index is greater than 1.96 the probability that the change in test score is due to random fluctuation in the measurement instrument is considered to be less than .05 and hence the change seen in the individual is considered reliable (from Jacobson & Truax, 1991).

\[
\frac{\text{POSTTEST SCORE} - \text{PRETEST SCORE}}{\text{STANDARD ERROR OF THE DIFFERENCE BETWEEN THE TWO TEST SCORES (S_{diff})}} = 1.96
\]

Therefore a change of greater than 12.01 points (i.e. 13 points) on the Total Problem Score can be considered reliable by Jacobson’s criteria, applied to the psychometric measures provided.
APPENDIX I

Details of statistical regression: predictors of outcome 6 months post-discharge

TWO STAGE MODEL REDUCTION STRATEGY: TIME 3 (6 MONTHS)
(See Section 5.5.1 of thesis).

STEP: STAGE 1(a): Run BIG (fully heterog) model:

\[
\text{MLR}_L \ dv = p_{3 \_tp} / \text{preds} = \begin{array}{c}
gender1 \ age1 \ p_{st1} \ ses1 \ p_{gf1} \ stp\_nat1 \\
oth\_nat1 \ jun\_mid1 \ sen\_mid1 \ los1 \ new\_th1 \\
alpha2 \ age2 \ p_{st2} \ ses2 \ p_{gf2} \ stp\_nat2 \\
oth\_nat2 \ jun\_mid2 \ sen\_mid2 \ los2 \ new\_th2 \\
alpha3 \ gender3 \ age3 \ p_{st3} \ ses3 \ p_{gf3} \ stp\_nat3 \\
oth\_nat3 \ jun\_mid3 \ sen\_mid3 \ los3 \ new\_th3 \\
alpha4 \ gender4 \ age4 \ p_{st4} \ ses4 \ p_{gf4} \ stp\_nat4 \\
oth\_nat4 \ jun\_mid4 \
\end{array}
\]

[NB. Note that ALPHA1 - 4 are collinear with (CONSTANT), so we must omit, say, ALPHA1. Also SEN\_MID4 has no variance; ALPHA2 and GENDER2 are perfectly correlated; so we omit ALPHA1, SEN\_MID4 and GENDER2 from analyses.]

Multiple R  .74337  
R Square .55260  
Adjusted R Square .32119  
Standard Error 21.87079

Analysis of Variance

<table>
<thead>
<tr>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>45</td>
<td>51400.8836</td>
</tr>
<tr>
<td>Residual</td>
<td>87</td>
<td>41614.8184</td>
</tr>
</tbody>
</table>

\[F = 2.38797 \text{ Signif F = .0003}\]

----------------- Variables in the Equation -----------------

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>Beta</th>
<th>T</th>
<th>Sig T (In group refinements immediately below)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER1</td>
<td>10.274202</td>
<td>11.365715</td>
<td>.151784</td>
<td>.904</td>
<td>.3685</td>
</tr>
<tr>
<td>AGE1</td>
<td>5.231034</td>
<td>3.787026</td>
<td>.858206</td>
<td>1.381</td>
<td>.1707</td>
</tr>
<tr>
<td>P_ST1</td>
<td>-4.837209</td>
<td>37.385620</td>
<td>-.044048</td>
<td>-.129</td>
<td>.8974</td>
</tr>
<tr>
<td>SES1</td>
<td>-1.06310</td>
<td>.259441</td>
<td>-.076811</td>
<td>.010</td>
<td>.9459</td>
</tr>
<tr>
<td>P_GF1</td>
<td>10.552682</td>
<td>14.761279</td>
<td>.350520</td>
<td>1.690</td>
<td>.0947</td>
</tr>
<tr>
<td>STP_Nat1</td>
<td>3.196983</td>
<td>18.947489</td>
<td>.01713</td>
<td>.169</td>
<td>.8664</td>
</tr>
<tr>
<td>OTH_Nat1</td>
<td>8.175765</td>
<td>10.240042</td>
<td>.088574</td>
<td>.798</td>
<td>.4268</td>
</tr>
<tr>
<td>JUN_Mid1</td>
<td>41.463690</td>
<td>17.368823</td>
<td>.465614</td>
<td>2.387</td>
<td>.0191</td>
</tr>
<tr>
<td>SEN_Mid1</td>
<td>-20.142734</td>
<td>15.797288</td>
<td>-.15085</td>
<td>-.127</td>
<td>.2057</td>
</tr>
<tr>
<td>LOS1</td>
<td>-.038486</td>
<td>1.423958</td>
<td>-.006299</td>
<td>-.027</td>
<td>.9785</td>
</tr>
<tr>
<td>NEW_Th1</td>
<td>-.395803</td>
<td>11.374736</td>
<td>-.077800</td>
<td>-.450</td>
<td>.6517</td>
</tr>
<tr>
<td>ALPHA2</td>
<td>147.466355</td>
<td>87.280628</td>
<td>2.357599</td>
<td>1.690</td>
<td>.0947</td>
</tr>
<tr>
<td>AGE2</td>
<td>.295710</td>
<td>4.343823</td>
<td>.049324</td>
<td>.068</td>
<td>.9459</td>
</tr>
<tr>
<td>P_ST2</td>
<td>18.128901</td>
<td>40.577086</td>
<td>.14313</td>
<td>.447</td>
<td>.6561</td>
</tr>
<tr>
<td>SES2</td>
<td>-.736928</td>
<td>.406023</td>
<td>-.567474</td>
<td>-.181</td>
<td>.0730</td>
</tr>
<tr>
<td>P_GF2</td>
<td>-17.431309</td>
<td>14.564541</td>
<td>-.569256</td>
<td>-.197</td>
<td>.2346</td>
</tr>
<tr>
<td>STP_Nat2</td>
<td>10.166809</td>
<td>16.934850</td>
<td>.057084</td>
<td>.600</td>
<td>.5498</td>
</tr>
</tbody>
</table>
STEP: STAGE 1 (b) for group 1:

** Drop those predictors in Group1 (****1s) with partial > 0.10 in BIG model:
** namely: all group 1 predictors but JUN_MID1:

```
MLR_L  dv = p_3_tp / preds =
   alpha2  age2  p_st2  ses2  p_gf2  stp_nat2
  oth_nat2  jun_mid2  sen_mid2  los2  new_th2
  p_st3  ses3  p_gf3  stp_nat3
  oth_nat3  jun_mid3  sen_mid3  los3  new_th3
  alpha4  gender4  age4  p_st4  ses4  p_gf4  stp_nat4
  oth_nat4  jun_mid4  los4  new_th4
```

Multiple R    .71463
R Square      .51070
Adjusted R Square  .33414
Standard Error 21.6617

Analysis of Variance

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>35</td>
<td>47502.71468</td>
<td>1357.22042</td>
</tr>
<tr>
<td>Residual</td>
<td>97</td>
<td>45512.99209</td>
<td>469.20610</td>
</tr>
</tbody>
</table>

F = 2.89259    Signif F = .0000
Increment test (vs BIG model):

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>SS</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIG model</td>
<td>51400.88836</td>
<td>45</td>
</tr>
<tr>
<td>BIG Residual</td>
<td>41614.81841</td>
<td>87</td>
</tr>
<tr>
<td>Reduced model</td>
<td>47502.71468</td>
<td>35</td>
</tr>
<tr>
<td>Increment</td>
<td>3898.17368</td>
<td>10</td>
</tr>
</tbody>
</table>

Increment-p = 0.615001

STEP: STAGE 1 (b) for group 2:

** Drop those ****2s (group 2) with partial > 0.10 in BIG:
** namely keep ALPHA2 SES2 OTH_NAT2:

\[
\text{MLR}_L \ dv = p_{3\_tp} / \text{preds} = \text{gender1 age1 p_st1 ses1 p_gf1 stp_nat1} \\
\text{oth_nat1 jun_mid1 sen_mid1 los1 new_th1} \\
\text{alpha2} \ 	ext{ses2} \\
\text{oth_nat2} \\
\text{alpha3 age3 p_st3 ses3 p_gf3 stp_nat3} \\
\text{oth_nat3 jun_mid3 sen_mid3 los3 new_th3} \\
\text{alpha4 gender4 age4 p_st4 ses4 p_gf4 stp_nat4} \\
\text{oth_nat4 jun_mid4} \\
\text{los4 new_th4} .
\]

Multiple R .72047
R Square .51907
Adjusted R Square .33176
Standard Error 21.69984

Analysis of Variance

<table>
<thead>
<tr>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>37</td>
<td>48281.80406</td>
</tr>
<tr>
<td>Residual</td>
<td>95</td>
<td>44733.90270</td>
</tr>
</tbody>
</table>

F = 2.77120 Signif F = .0000

Increment test (vs BIG model):

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>SS</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIG model</td>
<td>51400.88836</td>
<td>45</td>
</tr>
<tr>
<td>BIG Residual</td>
<td>41614.81841</td>
<td>87</td>
</tr>
<tr>
<td>Reduced model</td>
<td>48281.80406</td>
<td>37</td>
</tr>
<tr>
<td>Increment</td>
<td>3119.08430</td>
<td>8</td>
</tr>
</tbody>
</table>

Increment-p = 0.591245

STEP: STAGE 1(b) for group 3:

** Drop those ****3s (group 3) with partial > 0.10 in BIG:
** namely: keep ALPHA3 and STP_NAT3:

\[
\text{MLR}_L \ dv = p_{3\_tp} / \text{preds} = \text{gender1 age1 p_st1 ses1 p_gf1 stp_nat1} \\
\text{oth_nat1 jun_mid1 sen_mid1 los1 new_th1} \\
\text{alpha2} \ 	ext{age2 p_st2 ses2 p_gf2 stp_nat2} \\
\text{oth_nat2 jun_mid2 sen_mid2 los2 new_th2} \\
\text{alpha3} \ 	ext{stp_nat3} \\
\text{alpha4 gender4 age4 p_st4 ses4 p_gf4 stp_nat4} \\
\text{oth_nat4 jun_mid4} \ 	ext{los4 new_th4} .
\]
Multiple R           .68604
R Square             .47065
Adjusted R Square    .27964
Standard Error     22.53020

Analysis of Variance

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>35</td>
<td>43777.53066</td>
<td>1250.78659</td>
</tr>
<tr>
<td>Residual</td>
<td>97</td>
<td>49238.17610</td>
<td>507.61006</td>
</tr>
<tr>
<td>F</td>
<td>2.46407</td>
<td>Signif F = .0003</td>
<td></td>
</tr>
</tbody>
</table>

Increment test (vs BIG model):

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>SS</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIG model</td>
<td>51400.88836</td>
<td>45</td>
</tr>
<tr>
<td>BIG Residual</td>
<td>41614.81841</td>
<td>87</td>
</tr>
<tr>
<td>Reduced model</td>
<td>43777.53066</td>
<td>35</td>
</tr>
<tr>
<td>Increment</td>
<td>7623.3577</td>
<td>10</td>
</tr>
<tr>
<td>Increment-p</td>
<td>0.12182</td>
<td></td>
</tr>
</tbody>
</table>

STEP: STAGE 1(b) for group 4:

** Drop those ****4s (group 4) with partial > 0.10 in BIG:
** namely, keep AGE4, P_ST4, STP_NAT4, JUN_MID4, NEW_TH4.

MLR_L  dv = p_3_tp / preds = gender1 age1 p_st1 ses1 p_gf1 stp_nat1 oth_nat1 jun_mid1 sen_mid1 los1 new_th1 alpha2 age2 p_st2 ses2 p_gf2 stp_nat2 oth_nat2 jun_mid2 sen_mid2 los2 new_th2 alpha3 gender3 age3 p_st3 ses3 p_gf3 stp_nat3 oth_nat3 jun_mid3 sen_mid3 los3 new_th3 age4 p_st4 stp_nat4 jun_mid4 new_th4.

Multiple R           .72797
R Square             .52994
Adjusted R Square    .33282
Standard Error     21.68272

Analysis of Variance

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>39</td>
<td>49292.63873</td>
<td>1263.91381</td>
</tr>
<tr>
<td>Residual</td>
<td>93</td>
<td>43723.06804</td>
<td>470.14052</td>
</tr>
<tr>
<td>F</td>
<td>2.68837</td>
<td>Signif F = .0001</td>
<td></td>
</tr>
</tbody>
</table>

Increment test (vs BIG model):

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>SS</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIG model</td>
<td>51400.88836</td>
<td>45</td>
</tr>
<tr>
<td>BIG Residual</td>
<td>41614.81841</td>
<td>87</td>
</tr>
<tr>
<td>Reduced model</td>
<td>49292.63873</td>
<td>39</td>
</tr>
<tr>
<td>Increment</td>
<td>2108.24963</td>
<td>6</td>
</tr>
<tr>
<td>Increment-p</td>
<td>0.623091</td>
<td></td>
</tr>
</tbody>
</table>
STAGE 1 is now complete, and we admit the following predictors to STAGE 2:

- `jun_mid1`
- `alpha2 ses2 oth_nat2`
- `alpha3 stp_nat3`
- `age4 p_st4 stp_nat4 jun_mid4 new_th4`.

The model with these 11 predictors is called "PARTIAL1", and will be the base model in further increment tests aimed at arriving at a final model:

** Run model (call this model PARTIAL1):

```
MLR L  dv = p_3_tp / preds =
          jun_mid1 alpha2 ses2 oth_nat2
          alpha3 stp_nat3
          age4 p_st4 stp_nat4 jun_mid4 new_th4 .
```

---

Multiple R       .60340
R Square         .36409
Adjusted R Square .30628
Standard Error   22.10970

Analysis of Variance

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>11</td>
<td>33866.20193</td>
<td>3078.74563</td>
</tr>
<tr>
<td>Residual</td>
<td>121</td>
<td>59149.50484</td>
<td>488.83888</td>
</tr>
</tbody>
</table>

** F       = 6.29808  Signif F = .0000

------------------------ Variables in the Equation ------------------------

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>Beta</th>
<th>T</th>
<th>Sig T</th>
</tr>
</thead>
<tbody>
<tr>
<td>JUN_MID1</td>
<td>13.762471</td>
<td>7.852106</td>
<td>.154545</td>
<td>1.753</td>
<td>.0822</td>
</tr>
<tr>
<td>ALPHA2</td>
<td>18.238409</td>
<td>13.555706</td>
<td>.291584</td>
<td>1.345</td>
<td>.1810</td>
</tr>
<tr>
<td>SES2</td>
<td>-.428061</td>
<td>.300225</td>
<td>-.323821</td>
<td>-1.426</td>
<td>.1565</td>
</tr>
<tr>
<td>OTH_NAT2</td>
<td>26.327934</td>
<td>9.743475</td>
<td>.262522</td>
<td>2.702</td>
<td>.0079 KEEP</td>
</tr>
<tr>
<td>ALPHA3</td>
<td>33.522418</td>
<td>6.396021</td>
<td>.552953</td>
<td>5.241</td>
<td>.0000 KEEP</td>
</tr>
<tr>
<td>STP_NAT3</td>
<td>-20.313793</td>
<td>10.706273</td>
<td>-.146109</td>
<td>-1.897</td>
<td>.0602</td>
</tr>
<tr>
<td>AGE4</td>
<td>4.638348</td>
<td>1.158951</td>
<td>.820416</td>
<td>4.002</td>
<td>.0001 KEEP</td>
</tr>
<tr>
<td>P_ST4</td>
<td>-49.847150</td>
<td>23.255297</td>
<td>-.454989</td>
<td>-2.143</td>
<td>.0341 KEEP</td>
</tr>
<tr>
<td>STP_NAT4</td>
<td>-27.884355</td>
<td>13.721000</td>
<td>-.156563</td>
<td>-2.032</td>
<td>.0443 KEEP</td>
</tr>
<tr>
<td>JUN_MID4</td>
<td>24.325148</td>
<td>10.884563</td>
<td>.205393</td>
<td>2.235</td>
<td>.0273 KEEP</td>
</tr>
<tr>
<td>NEW_TH4</td>
<td>19.911851</td>
<td>8.865828</td>
<td>.215921</td>
<td>2.246</td>
<td>.0265 KEEP</td>
</tr>
<tr>
<td>(Constant)</td>
<td>31.391375</td>
<td>4.904342</td>
<td>.769218</td>
<td>6.401</td>
<td>.0000</td>
</tr>
</tbody>
</table>

---

STEP STAGE 2(a): Omit predictors with partial p > 0.05:

** Reduce again using increment-p of 0.10 (when DF > 1; call this model PARTIAL2):

```
MLR L  dv = p_3_tp / preds = oth_nat2
          alpha3
          age4 p_st4 stp_nat4 jun_mid4 new_th4 .
```

---

Multiple R       .56246
R Square         .31636
Adjusted R Square .27808
Standard Error   22.55469

Analysis of Variance

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>7</td>
<td>29426.43472</td>
<td>4203.77639</td>
</tr>
<tr>
<td>Residual</td>
<td>125</td>
<td>63589.27205</td>
<td>508.71418</td>
</tr>
</tbody>
</table>

** F       = 8.26353  Signif F = .0000
Increment test (vs PARTIAL1 model):

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>SS</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARTIAL1 model</td>
<td>33866.20193</td>
<td>11</td>
</tr>
<tr>
<td>PARTIAL1 Res.</td>
<td>59149.50484</td>
<td>121</td>
</tr>
<tr>
<td>Reduced model</td>
<td>29426.43472</td>
<td>7</td>
</tr>
<tr>
<td>Increment</td>
<td>4439.76721</td>
<td>4</td>
</tr>
<tr>
<td>Increment-p</td>
<td>0.065513</td>
<td></td>
</tr>
</tbody>
</table>

STEP: STAGE 2(b): As increment-p < 0.10, re-admit "best" omitted term from PARTIAL1 - viz. STP_NAT3.

MLR_L dv = p_3_tp / preds = oth_nat2
  alpha3 stp_nat3
  age4 p_st4 stp_nat4 jun_mid4 new_th4.

Multiple R           .57903
R Square              .33528
Adjusted R Square     .29239
Standard Error       22.32990

Analysis of Variance
   DF      Sum of Squares      Mean Square
Regression          8         31186.26641       3898.28330
Residual            124        61829.44035        498.62452
F =       7.81807       Signif F =  .0000

Increment test (vs PARTIAL1 model):

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>SS</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARTIAL1 model</td>
<td>33866.20193</td>
<td>11</td>
</tr>
<tr>
<td>PARTIAL1 Res.</td>
<td>59149.50484</td>
<td>121</td>
</tr>
<tr>
<td>Reduced model</td>
<td>31186.26641</td>
<td>8</td>
</tr>
<tr>
<td>Increment</td>
<td>2679.93552</td>
<td>3</td>
</tr>
<tr>
<td>Increment-p</td>
<td>0.145764</td>
<td></td>
</tr>
</tbody>
</table>

As this increment test is OK, we now have our final model:

MLR_L dv = p_3_tp / preds = oth_nat2
  alpha3 stp_nat3
  age4 p_st4 stp_nat4 jun_mid4 new_th4.

Multiple R           .57903
R Square              .33528
Adjusted R Square     .29239
Standard Error       22.32990

Analysis of Variance
   DF      Sum of Squares      Mean Square
Regression          8         31186.26641       3898.28330
Residual            124        61829.44035        498.62452
F =       7.81807       Signif F =  .0000
<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>Beta</th>
<th>T</th>
<th>Sig T</th>
</tr>
</thead>
<tbody>
<tr>
<td>OTH_NAT2</td>
<td>17.205240</td>
<td>7.683828</td>
<td>.171530</td>
<td>2.239</td>
<td>.0270</td>
</tr>
<tr>
<td>ALPHA3</td>
<td>29.616333</td>
<td>5.135436</td>
<td>.488522</td>
<td>5.767</td>
<td>.0000</td>
</tr>
<tr>
<td>STP_NAT3</td>
<td>-20.313793</td>
<td>10.812902</td>
<td>-.146109</td>
<td>-1.879</td>
<td>.0626</td>
</tr>
<tr>
<td>AGE4</td>
<td>4.342545</td>
<td>1.132252</td>
<td>.768095</td>
<td>3.835</td>
<td>.0002</td>
</tr>
<tr>
<td>P_ST4</td>
<td>-51.027898</td>
<td>23.457019</td>
<td>-.465766</td>
<td>-2.175</td>
<td>.0315</td>
</tr>
<tr>
<td>STP_NAT4</td>
<td>-28.026416</td>
<td>13.856921</td>
<td>-.157361</td>
<td>-2.023</td>
<td>.0453</td>
</tr>
<tr>
<td>JUN_MID4</td>
<td>22.712663</td>
<td>10.873296</td>
<td>.191778</td>
<td>2.089</td>
<td>.0388</td>
</tr>
<tr>
<td>NEW_TH4</td>
<td>20.202778</td>
<td>8.949369</td>
<td>.219075</td>
<td>2.257</td>
<td>.0257</td>
</tr>
<tr>
<td>(Constant)</td>
<td>35.297460</td>
<td>3.029646</td>
<td></td>
<td>11.651</td>
<td>.0000</td>
</tr>
</tbody>
</table>
### APPENDIX J

**Details of statistical regression: predictors of outcome 12 months post-discharge**

**TWO STAGE MODEL REDUCTION STRATEGY : TIME 4 (12 MONTHS)**  
(See Section 5.5.1 of thesis).

**STEP: STAGE 1(a): Run BIG (fully heterog) model:**

```plaintext
MLR_L dv = p_3_tp / preds = gender1 age1 p_st1 ses1 p_gf1 stp_nat1 oth_nat1 jun_mid1 sen_mid1 los1 new_th1 alpha2 age2 p_st2 ses2 p_gf2 stp_nat2 oth_nat2 jun_mid2 sen_mid2 los2 new_th2 alpha3 age3 p_st3 ses3 p_gf3 stp_nat3 oth_nat3 jun_mid3 sen_mid3 los3 new_th3 alpha4 age4 p_st4 ses4 p_gf4 stp_nat4 oth_nat4 jun_mid4 los4 new_th4 .
```

[NB. Note that ALPHA1-4 are collinear with (CONSTANT), so we must omit, say, ALPHA1. Also SEN_MID4 has no variance; ALPHA2 and GENDER2 are perfectly correlated; so we omit ALPHA1, SEN-MID4 and GENDER2 from analyses.]

**Multiple R** .76717  
**R Square** .58854  
**Adjusted R Square** .37814  
**Standard Error** 22.31813

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>45</td>
<td>62697.65610</td>
<td>1393.28125</td>
</tr>
<tr>
<td>Residual</td>
<td>88</td>
<td>43832.7103</td>
<td>498.09906</td>
</tr>
</tbody>
</table>

**F =** 2.79720  
**Signif F = .0000**

**------------------------ Variables in the Equation ------------------------**

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>Beta</th>
<th>T</th>
<th>Sig T</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER1</td>
<td>17.094735</td>
<td>11.587264</td>
<td>.236187</td>
<td>1.475</td>
<td>.1437</td>
</tr>
<tr>
<td>AGE1</td>
<td>3.857883</td>
<td>3.482538</td>
<td>.613970</td>
<td>1.108</td>
<td>.2710</td>
</tr>
<tr>
<td>P_ST1</td>
<td>-31.618169</td>
<td>38.358423</td>
<td>-.274449</td>
<td>-.824</td>
<td>.4120</td>
</tr>
<tr>
<td>SES1</td>
<td>-2.17689</td>
<td>.277374</td>
<td>-.149863</td>
<td>-.785</td>
<td>.4347</td>
</tr>
<tr>
<td>P_GF1</td>
<td>14.017229</td>
<td>45.145583</td>
<td>.453645</td>
<td>1.059</td>
<td>.2925</td>
</tr>
<tr>
<td>STP_NAT1</td>
<td>-1.409036</td>
<td>25.587669</td>
<td>-.004301</td>
<td>-.055</td>
<td>.9562</td>
</tr>
<tr>
<td>OTH_NAT1</td>
<td>7.69721</td>
<td>10.049994</td>
<td>.083497</td>
<td>.766</td>
<td>.4558</td>
</tr>
<tr>
<td>JUN_MID1</td>
<td>35.528496</td>
<td>19.794084</td>
<td>.397289</td>
<td>1.795</td>
<td>.0761</td>
</tr>
<tr>
<td>SEN_MID1</td>
<td>-18.114010</td>
<td>16.025082</td>
<td>-.182497</td>
<td>-.1130</td>
<td>.2614</td>
</tr>
<tr>
<td>LOS1</td>
<td>-96.4284</td>
<td>12.71605</td>
<td>-.143327</td>
<td>.758</td>
<td>.4503</td>
</tr>
<tr>
<td>NEW_TH1</td>
<td>-3.812388</td>
<td>12.085065</td>
<td>-.040159</td>
<td>.315</td>
<td>.7532</td>
</tr>
</tbody>
</table>

**ALPHA2**  
<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>Beta</th>
<th>T</th>
<th>Sig T</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER2</td>
<td>11.081427</td>
<td>101.305762</td>
<td>.169326</td>
<td>.132</td>
<td>.9596</td>
</tr>
<tr>
<td>AGE2</td>
<td>5.463234</td>
<td>3.465443</td>
<td>.848470</td>
<td>1.576</td>
<td>.1185</td>
</tr>
<tr>
<td>P_ST2</td>
<td>-15.7625</td>
<td>44.145583</td>
<td>-.001348</td>
<td>.004</td>
<td>.9972</td>
</tr>
<tr>
<td>SES2</td>
<td>-3.49147</td>
<td>3.51446</td>
<td>-.269019</td>
<td>.993</td>
<td>.3232</td>
</tr>
<tr>
<td>P_GF2</td>
<td>5.516144</td>
<td>13.990760</td>
<td>.170665</td>
<td>.394</td>
<td>.6943</td>
</tr>
<tr>
<td>STP_NAT2</td>
<td>29.310503</td>
<td>17.149971</td>
<td>.153791</td>
<td>1.709</td>
<td>.0910</td>
</tr>
<tr>
<td>OTH_NAT2</td>
<td>21.378037</td>
<td>12.901372</td>
<td>.208127</td>
<td>1.657</td>
<td>.1011</td>
</tr>
<tr>
<td>JUN_MID2</td>
<td>16.944722</td>
<td>20.233206</td>
<td>.177873</td>
<td>.837</td>
<td>.4046</td>
</tr>
<tr>
<td>SEN_MID2</td>
<td>-39.871390</td>
<td>18.984410</td>
<td>-.209203</td>
<td>-.2100</td>
<td>.0386</td>
</tr>
<tr>
<td>LOS2</td>
<td>-97.7743</td>
<td>1.030448</td>
<td>-.17889</td>
<td>.95</td>
<td>.9246</td>
</tr>
<tr>
<td>NEW_TH2</td>
<td>-6.718750</td>
<td>9.631531</td>
<td>-.076890</td>
<td>.698</td>
<td>.4873</td>
</tr>
</tbody>
</table>

**ALPHA3**  
<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>Beta</th>
<th>T</th>
<th>Sig T</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER3</td>
<td>100.671172</td>
<td>101.305762</td>
<td>1.470313</td>
<td>.994</td>
<td>.3231</td>
</tr>
<tr>
<td>AGE3</td>
<td>-25.522586</td>
<td>16.525788</td>
<td>-.352630</td>
<td>-.154</td>
<td>.1261</td>
</tr>
<tr>
<td>P_ST3</td>
<td>-11.327275</td>
<td>45.119381</td>
<td>-.096718</td>
<td>-.251</td>
<td>.8024</td>
</tr>
</tbody>
</table>

[See Section 5.5.1 of thesis.]
**DROP those predictors in Group1 (**1s) with partial > 0.10 in BIG model:
** namely: all group 1 predictors but JUN_MID1:

**STEP: STAGE 1 (b) for group 1:**

```
MLR_L  dv = p_3_tp / preds =

jun_mid1 alpha2 age2 p_st2 ses2 p_gf2 stp_nat2
oth_nat2 jun_mid2 sen_mid2 los2 new_th2
alpha3 gender3 age3 p_st3 ses3 p_gf3 stp_nat3
oth_nat3 jun_mid3 sen_mid3 los3 new_th3
alpha4 gender4 age4 p_st4 ses4 p_gf4 stp_nat4
oth Nat4 jun_mid4

Multiple R .74175
R Square .55019
Adjusted R Square .38954
Standard Error 22.11251
```

Analysis of Variance

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>35</td>
<td>58611.98241</td>
<td>1674.62807</td>
</tr>
<tr>
<td>Residual</td>
<td>98</td>
<td>47918.39073</td>
<td>488.96317</td>
</tr>
</tbody>
</table>

F = 3.42486    Signif F = .0000

Increment test (vs BIG model):

```
EFFECT   SS
BIG model 62697.65610 45
BIG Residual 43832.71703 88
Reduced model 58611.98241 35
Increment 4085.67369 10
Increment-p = 0.610049
```

**STEP: STAGE 1 (b) for group 2:**

**DROP those (**2s) (group 2) with partial > 0.10 in BIG:
** namely keep STP_NAT2 and SEN_MID2:
MLR_L  dv = p_4_tp / preds =  gender1 age1 p_st1 ses1 p_gf1 stp_nat1
oth_nat1 jun_mid1 sen_mid1 los1 new_th1
              stp_nat2
          sen_mid2
          alpha3 gender3 age3 p_st3 ses3 p_gf3 stp_nat3
oth_nat3 jun_mid3 sen_mid3 los3 new_th3
oth_nat4 jun_mid4
Multiple R   .73151
R Square      .53510
Adjusted R Square  .36256
Standard Error   22.59591

Analysis of Variance

<table>
<thead>
<tr>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Regression</td>
<td>57004.56378</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>49525.80935</td>
</tr>
</tbody>
</table>

F = 3.10132  Signif F = .0000

Increment test (vs BIG model):

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>SS</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIG model</td>
<td>62697.65610</td>
<td>45</td>
</tr>
<tr>
<td>BIG Residual</td>
<td>43832.71703</td>
<td>88</td>
</tr>
<tr>
<td>Reduced model</td>
<td>57004.56378</td>
<td>36</td>
</tr>
</tbody>
</table>

Increment: 5693.56378  9  Increment-p = 0.264585

STEP: STAGE 1(b) for group 3:

** Drop those ****3s (group 3) with partial > 0.10 in BIG:
** namely: keep SES3 and LOS3:

MLR_L  dv = p_4_tp / preds =  gender1 age1 p_st1 ses1 p_gf1 stp_nat1
oth_nat1 jun_mid1 sen_mid1 los1 new_th1
              stp_nat2
          sen_mid2
          alpha3 gender3 age3 p_st3 ses3 p_gf3 stp_nat3
oth_nat3 jun_mid3 sen_mid3 los3 new_th3
alpha4 gender4 age4 p_st4 ses4 p_gf4 stp_nat4
oth_nat4 jun_mid4
Multiple R   .72757
R Square      .52935
Adjusted R Square  .36126
Standard Error   22.61894

Analysis of Variance

<table>
<thead>
<tr>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Regression</td>
<td>56391.97797</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>50138.39517</td>
</tr>
</tbody>
</table>

F = 3.14923  Signif F = .0000
Increment test (vs BIG model):

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>SS</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIG model</td>
<td>62697.65610</td>
<td>45</td>
</tr>
<tr>
<td>BIG Residual</td>
<td>43832.71703</td>
<td>88</td>
</tr>
<tr>
<td>Reduced model</td>
<td>56391.9797</td>
<td>35</td>
</tr>
</tbody>
</table>

| Increment    | 6305.67813    | 10 |

Increment-p = 0.262092

**STEP: STAGE 1(b) for group 4:**

** Drop those ****4s (group 4) with partial > 0.10 in BIG: 
** namely, keep AGE4, SES4, STP_NAT4, NEW_TH4 

MLR_L  dv = p_4_tp / preds = gender1 age1 p_st1 ses1 p_gf1 stp_nat1 oth_nat1 jun_mid1 sen_mid1 los1 new_th1
        alpha2         age2 p_st2 ses2 p_gf2 stp_nat2 oth_nat2 jun_mid2 sen_mid2 los2 new_th2
        alpha3 gender3 age3 p_st3 ses3 p_gf3 stp_nat3 oth_nat3 jun_mid3 sen_mid3 los3 new_th3
        age4 ses4 stp_nat4 new_th4.

Multiple R           .74294
R Square             .55196
Adjusted R Square    .37274
Standard Error     22.41477

Analysis of Variance

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>38</td>
<td>58800.29103</td>
<td>1547.37608</td>
</tr>
<tr>
<td>Residual</td>
<td>95</td>
<td>47730.08211</td>
<td>502.42192</td>
</tr>
</tbody>
</table>

F = 3.07983   Signif F = .0000

Increment test (vs BIG model):

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>SS</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIG model</td>
<td>62697.65610</td>
<td>45</td>
</tr>
<tr>
<td>BIG Residual</td>
<td>43832.71703</td>
<td>88</td>
</tr>
<tr>
<td>Reduced model</td>
<td>58800.29103</td>
<td>38</td>
</tr>
</tbody>
</table>

| Increment    | 3897.36507    | 7  |

Increment-p = 0.359265

**STAGE 1 is now complete, and we admit the following predictors to STAGE 2:**

jun_mid1
stp_nat2 sen_mid2
ses3 los3
age4 ses4 stp_nat4 new_th4.

The model with these 11 predictors is called "PARTIAL1", and will be the base model in further increment tests aimed at arriving at a final model:

** Run model (call this model PARTIAL1):

MLR_L  dv = p_4_tp / preds =jun_mid1
        stp_nat2 sen_mid2
        ses3 los3
        age4 ses4 stp_nat4 new_th4.
Multiple R   .63467
R Square     .40280
Adjusted R Square .35946
Standard Error  22.65087

Analysis of Variance

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>42910.71169</td>
<td>9</td>
<td>4767.85685</td>
</tr>
<tr>
<td>Residual</td>
<td>63619.66145</td>
<td>124</td>
<td>513.06179</td>
</tr>
</tbody>
</table>

F = 9.29295  Signif F = .0000

---------------------- Variables in the Equation ----------------------

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>Beta</th>
<th>T</th>
<th>Sig T</th>
</tr>
</thead>
<tbody>
<tr>
<td>JUN_MID1</td>
<td>1.907341</td>
<td>6.682375</td>
<td>.021328</td>
<td>.285</td>
<td>.7758</td>
</tr>
<tr>
<td>STP_NAT2</td>
<td>23.007341</td>
<td>13.471103</td>
<td>.120718</td>
<td>1.708</td>
<td>.0902</td>
</tr>
<tr>
<td>SEN_MID2</td>
<td>.840675</td>
<td>13.471103</td>
<td>.004411</td>
<td>.062</td>
<td>.9503</td>
</tr>
<tr>
<td>SES3</td>
<td>.294207</td>
<td>.169681</td>
<td>.208032</td>
<td>1.734</td>
<td>.0854</td>
</tr>
<tr>
<td>LOS3</td>
<td>1.279533</td>
<td>.828962</td>
<td>.182958</td>
<td>1.544</td>
<td>.1252</td>
</tr>
<tr>
<td>AGE4</td>
<td>.878291</td>
<td>.926385</td>
<td>.149256</td>
<td>.948</td>
<td>.3449</td>
</tr>
<tr>
<td>SES4</td>
<td>.480383</td>
<td>.196751</td>
<td>.381205</td>
<td>2.442</td>
<td>.0160</td>
</tr>
<tr>
<td>STP_NAT4</td>
<td>-38.704199</td>
<td>12.488280</td>
<td>-.233599</td>
<td>3.099</td>
<td>.0024</td>
</tr>
<tr>
<td>NEW_TH4</td>
<td>23.030905</td>
<td>7.905239</td>
<td>.250701</td>
<td>2.913</td>
<td>.0042</td>
</tr>
</tbody>
</table>

(Constant) | 33.492659 | 3.232648 | 10.361 | .0000 |

STEP STAGE 2(a): Omit predictors with partial p > 0.05:

** Reduce again using increment-p of 0.10 (when DF > 1; call this model PARTIAL2):

MLR_L  dv = p_4_tp / preds = ses4 stp_nat4 new_th4 .

Multiple R   .52551
R Square     .27616
Adjusted R Square .25945
Standard Error  24.35495

Analysis of Variance

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>29419.08495</td>
<td>3</td>
<td>9806.36165</td>
</tr>
<tr>
<td>Residual</td>
<td>77111.28818</td>
<td>130</td>
<td>593.16376</td>
</tr>
</tbody>
</table>

F = 16.53230  Signif F = .0000

Increment test (vs PARTIAL1 model):

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>SS</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARTIAL1 model</td>
<td>42910.71169</td>
<td>9</td>
</tr>
<tr>
<td>PARTIAL1 Res</td>
<td>63619.66145</td>
<td>124</td>
</tr>
<tr>
<td>Reduced model</td>
<td>29419.08495</td>
<td>3</td>
</tr>
</tbody>
</table>

Increment | 13491.62674| 6 |

Increment-p  =  0.000475
**STEP: STAGE 2(b):** As increment-\( p < 0.10 \), re-admit "best" omitted term from PARTIAL1 - viz. LOS3. 

\[
\text{MLR}_L \ dv = p_{4\_tp} / \text{preds} = \text{los3 ses4 stp\_nat4 new\_th4}.
\]

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>Beta</th>
<th>T</th>
<th>Sig T</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOS3</td>
<td>2.258867</td>
<td>.506759</td>
<td>.322992</td>
<td>4.457</td>
<td>.0000</td>
</tr>
<tr>
<td>SES4</td>
<td>.591050</td>
<td>.112162</td>
<td>.469025</td>
<td>5.270</td>
<td>.0000</td>
</tr>
<tr>
<td>STP_NAT4</td>
<td>-39.072824</td>
<td>12.507458</td>
<td>-0.235824</td>
<td>-3.124</td>
<td>.0022</td>
</tr>
<tr>
<td>NEW_TH4</td>
<td>24.422788</td>
<td>7.725563</td>
<td>.265852</td>
<td>3.161</td>
<td>.0020</td>
</tr>
<tr>
<td>(Constant)</td>
<td>36.254755</td>
<td>2.611246</td>
<td>13.884</td>
<td>.0000</td>
<td></td>
</tr>
</tbody>
</table>

Increment test (vs PARTIAL1 model):

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>SS</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARTIAL1 model</td>
<td>42910.71169</td>
<td>9</td>
</tr>
<tr>
<td>PARTIAL1 Res</td>
<td>63619.66145</td>
<td>124</td>
</tr>
<tr>
<td>Reduced model</td>
<td>39710.89498</td>
<td>4</td>
</tr>
<tr>
<td>Increment</td>
<td>3199.81671</td>
<td>5</td>
</tr>
</tbody>
</table>

Increment-\( p = 0.291156 \)

As this increment test is OK, we now have our final model:

\[
\text{MLR}_L \ dv = p_{4\_tp} / \text{preds} = \text{los3 ses4 stp\_nat4 new\_th4}.
\]

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>Beta</th>
<th>T</th>
<th>Sig T</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOS3</td>
<td>2.258867</td>
<td>.506759</td>
<td>.322992</td>
<td>4.457</td>
<td>.0000</td>
</tr>
<tr>
<td>SES4</td>
<td>.591050</td>
<td>.112162</td>
<td>.469025</td>
<td>5.270</td>
<td>.0000</td>
</tr>
<tr>
<td>STP_NAT4</td>
<td>-39.072824</td>
<td>12.507458</td>
<td>-0.235824</td>
<td>-3.124</td>
<td>.0022</td>
</tr>
<tr>
<td>NEW_TH4</td>
<td>24.422788</td>
<td>7.725563</td>
<td>.265852</td>
<td>3.161</td>
<td>.0020</td>
</tr>
<tr>
<td>(Constant)</td>
<td>36.254755</td>
<td>2.611246</td>
<td>13.884</td>
<td>.0000</td>
<td></td>
</tr>
</tbody>
</table>