

**B4 School Report:  
A Critique of a Child Health Screening  
and Intervention Programme**

**Robert Miller**  
(finalised: 6.12.2013)

## Contents:

<b>SUMMARY</b> .....	4
<b>A: Introduction: Brief History of the Background to     <i>B4 School Check</i></b> .....	6
<b>B: What has Driven the Mental Health Components of the     <i>B4 School Check?</i></b> .....	9
<b>C: How the Objectives of <i>B4 School Check</i> are Presented</b> .....	11
[I] <i>History of Screening for Behavioural/Emotional         Difficulties in Pre-Schoolers and Children</i> .....	11
[II] <i>Apparent Objectives of Behavioural/Developmental         Component of B4 School Go Beyond Those Advised in its         Supporting Documents</i> .....	11
[III] <i>Disconnection Between Different Parts of Ministry         of Health’s Agenda for Child and Youth Mental Health</i> .....	12
[IV] <i>Disconnection Between Aims and Methods of         Behavioural Testing in B4 School, and Supporting Documents</i> ....	13
[V] <i>Which Aspects of Child Mental Health Does <b>B4 School</b>         Screen For?</i> .....	14
[VI] <i>Roots of Anti-Social Behaviour in Children: An Individual         or a Societal Problem?</i> .....	15
[VII] <i>A Hidden Agenda?</i> .....	17
<b>D: What Were the Intended Ways of Implementing the     SDQ, and of Using Information Derived From It?</b> .....	19
[I] <i>By Whom, and How is Rating via SDQ conducted?</i> .....	19
[II] <i>Advice to Practitioners to Use of Language to Hide         the Real Objectives of the SDQ: Issue of Truly Informed         Consent.</i> ....	20
[III] <i>Referral Criteria and Referral Streams</i> .....	21
<b>E: Critique of the SDQ</b> .....	22
[I] <i>Introduction</i> .....	22
[II] <i>The Overall Style and Mode of Delivery of the SDQ</i> .....	23

[III] <i>Validation of the SDQ in the Target Age-group</i> .....	25
[IV] <i>Biased Coverage of Child Mental Health Problems by the SDQ</i> .....	25
[V] <i>Ability of the SDQ to Discriminate Different Types of Problem</i> .....	29
[VI] <i>Specificity, and the Problem of False Positives</i> .....	31
[VII] <i>Comparison with SDQ: Merits of Screening Methods to Detect or Identify Problems, in Other Sections of B4 School, and in Related Fields</i> .....	38
<b>F: Evidence on Effectiveness of Intervention in Childhood Years as a Public Health/Social Development Measure</b> .....	<b>51</b>
<b>G: How have the SDQ, and the Data it Gives Been Used?</b> .....	<b>60</b>
[I] <i>Hawke’s Bay Report, 2010</i> .....	60
[II] <i>Editorial in NZMJ, 2010</i> .....	61
[III] <i>Two Recent Sources of Data</i> .....	61
[IV] <i>Policy Implications</i> .....	63
<b>H: Implications for Australia?</b> .....	<b>63</b>
<b>I: Recommendations</b> .....	<b>64</b>

## SUMMARY:

(i) Starting in late 2008, a programme for pre-school health checking has been rolled out for 4-5 year-olds, across New Zealand. Most of it is standard practice in many countries, using precise, well-tried tests, with clear referral pathways when an abnormality is detected. An area of concern, on which this report focuses is screening for behavioural and emotional difficulties, using an instrument called the “Strengths and Difficulties Questionnaire” (SDQ), developed over ten years ago at the Institute of Psychiatry in London. The reasons for concern are as follows:

(ii) The motives behind use of the SDQ within *B4 School* are not transparent. What is presented as a screening tool for incipient mental health issues appears, on closer examination, to be motivated mainly by an attempt to identify at an early age youngsters with potential in later years for disruptive behaviour, delinquency as teenagers, and, as adults, criminality.

(iii) The SDQ was designed for use with 4-17-year olds. It has never been validated specifically for the 4-5-year age group. As such, its first deployment should have been declared as a small-scale research project for this age group with appropriate ethical scrutiny, not as a mass programme intended for near-universal screening of the age group.

(iv) In psychiatry, it may be possible to detect that a person has a problem in the mental health long before the nature of the problem can be identified. As presented to parents and teachers, the SDQ claims not only to detect problems, but also, to some extent, to identify several broad categories of problem. However, its ability to discriminate problems is very limited, so that the SDQ, as presented, lays claim to a degree of scientific authority and precision which cannot be supported.

(v) The SDQ, at its best, may be useful for detecting externalising disorders (DSM-IV categories such as Conduct Disorder, Oppositional-Defiant Disorder, and ADHD), but is not the best instrument for screening for a full range of commonly-occurring psychological problems of children. Even for the disorders for which it is best suited (and for “any problem” determinations), it has an excess of false positive detections. This is worrying when services are often inadequate, even now, to deliver good quality treatment when serious problems are identified.

(vi) The problem of the high rate of false positives using the SDQ raises ethical concerns. Given the personal sensitivity of issues about mental health data, the history of abuse in state mental health facilities in New Zealand, and the lack of ethical scrutiny of the *B4 School Check* (and therefore the lack of transparency on the fate of data collected, and its possible future uses), this is a cause for serious concern.

(vi) While parents are required to sign a form giving informed consent, the clear intention (reinforced by the setting of performance targets for screening by DHBs, and alliances with TV campaigns aimed at pre-schoolers etc), is

that it be used in near-universally across the age group. This is inconsistent with ethical requirements for truly informed consent.

(vii) The referral processes for children found “borderline” or “abnormal” by the SDQ are not clear, in contravention of guidelines for health screening programmes.

(viii) Many studies have examined intervention aiming to reduce adverse outcomes for youngsters identified as having Conduct Disorder and related problems. Although they usually show that effective intervention is possible in principle, none has yet shown that this effectiveness can be transferred to a population-wide preventive program of social/psychological medicine. While research studies should continue, great caution is needed in implementing findings currently available, especially if this is to be done nation-wide.

(ix) Data on use of the SDQ to date within the *B4 School Check*, indicate reluctance to use it (especially by teachers), large differences between DHBs in the proportion of children screened with the SDQ and found to be “borderline” or “abnormal”, and truly massive differences between DHBs on the rates of referral for children in these categories. There are indications that in most DHBs, services could not cope with demand if all in these categories were to be referred. While Ministry of Health collects data on the number of children screened using the SDQ, the Ministry appears not to be collecting data on outcomes as a result of referral, which is inconsistent with current policy emphases.

(x) Guidelines issued for Health Screening Programmes (2003), are not met on most counts, for use of the SDQ within *B4 School*.

(xi) Serious issues are raised about the over-enthusiastic approach to early intervention, poor understanding of scientific methodology, use of quasi-medical approaches within a field which is more properly seen in a different context, as well as the governance structure and accountability systems behind the introduction of the SDQ into the *B4 School Check*.

(xii) Little consideration is given to the roots of antisocial behaviour at a social level. While the importance of intervention at an individual level is not denied, social policy should also address wider social roots of antisocial behaviour.

(xiii) This report is not an attack on preventive medicine generally, nor in psychiatry (where it is certainly a relative newcomer). It is also not an attack on the idea of exploring preventive measures to reduce rates of transition from conduct disorder in children to more serious problems in adolescence and adulthood. However, strategies which are developed should be more self-critical about the scientific merits of any system, their ethical probity, their cost-benefit balance, and whether preventive measures should best be implemented by mass screening and intervention, screening and intervention targeted at selected groups, or ones closely integrated with regular clinical services.

# **B4 School Report: A Critique of a Child Health Screening and Intervention Programme**

## **A: Introduction: Brief History of the Background to *B4 School Health Check*.**

In the year 2009 a health check system was introduced in New Zealand for pre-schoolers (age 4-5 years), entitled *B4 School*. The objectives of the *B4 School Health Check* are described in the *Practitioners Handbook*, for use by health practitioners who supervise the implementation of this health check:

“The Ministry of Health surveyed providers of the [previous] School New Entrant Check and found it was not being provided universally, and that only a few children were receiving a full assessment. In some areas, the School New Entrant Check was being done by several different people among whom there appeared to be little communication”.

Looking further back, the context for this remark is probably that, in the mid-1990s the number of practice nurses, working to assess new school entrants was reduced. They were often trained Public Health nurses or Plunkett nurses, with experience and expertise in child health. The reduction was a probable cause of the short-fall in completions using the New Entrant Check. Clearly the intention is that *this* health check should be rolled out across the country, to allow near-universal screening/assessment for remediable health problems. It is then hoped that, if problems are identified early, they can be corrected; and it is envisaged that large benefits will follow, for the children, their families and the health system generally.

I write this report because in New Zealand I have heard concern expressed about the mental health component of *B4 School*. In addition, there is an intention to implement something similar across Australia, drawing partly on New Zealand experience with this component of pre-school health screening. The issue is thus bi-national. A recent visitor to both countries (Professor Allen Frances, from Duke University, North Carolina, who headed the committees which assembled the American Psychiatric Association’s *DSM-IV*) has also expressed concern that this sort of screening contributes to what he calls “diagnostic inflation” in psychiatry. My own assessment of *B4 School*, and its mental health component is based not on any opinions I have heard, but on my own reading about the background and ways of promoting *B4 School*, its methodology, and how data so collected are being used for referral or other outcomes. Health screening programs usually are (and should be) matters for public scrutiny, debate, and sometimes controversy. This report can be seen as a contribution to the debate on *B4 School*. In several of its later sections covering details of *B4 School*, wider policy issues are raised, and also form part of this debate. I remind readers that I sit on committees of the Royal Australia and New Zealand College of Psychiatrists, as a community representative; and since the

issues discussed in this report are bi-national, it is of relevance to a bi-national college such as RANZCP. Nevertheless the report was not commissioned from within RANZCP committees, so it can be thought of as a fully independent and public document.

There are several sections in the *B4 School Check*. Each section includes details of appropriate referral pathways. Some parts - where research data is insufficient to guide referral - are more about population surveillance than individual guidance, to provide information on the basis of which strategies of action can be taken at some future time, perhaps at a population rather than an individual level. The sections are as follows:

- (i) *General Health*. This takes up a single page in the *Practitioner's Handbook*. It asks for factual information about immunization, and provides an opportunity to give scheduled immunizations which have been missed. It also asks about on-going treatment for eye or ear problems.
- (ii) *Hearing Screening* (8 pages in the *Handbook*). This involves audiometry tests ("sweep test") and, for infants under 6 months, auditory evoked potentials. Audiometry screening is based on guidelines for screening from the American Speech-Language-Hearing Association<sup>1</sup>. When a "sweep test" is abnormal, tympanometry is to be carried out and followed up along a clinical referral pathway. "If sweep test is abnormal and tympanometry is normal, the child may have a sensorineural hearing loss. In this situation refer the child to audiology, or ear, nose and throat services (depending on local pathways for assessment) . . . If tympanogram is abnormal, refer to GP or ear nurse for assessment."
- (iii) *Vision Screening* (7 pages in *Handbook*). This includes "4-m Parr charts", with and without confusion bars and "4-m Snellen charts". Referral guidelines include the possibility of re-screening after an interval of 3-6 months, or referral to an ophthalmic specialist. There are also guidelines for action after re-screening. However the comment is also made that "protocol for vision screening referral is being developed. Until then, local protocol should be used".
- (iv) *Oral Health Screening and Promotion* (2 pages in *Handbook*). Details of the System of screening is to be found in *Healthy Smile, Healthy Child* available from the New Zealand Dental Association.
- (v) *Identifying Behavioural Problems* (8.5 pages in *Handbook*), and *Developmental Screening and Surveillance* (6.5 pages in *Handbook*). The system of screening for *Behavioural Problems* is the main focus of the present report. *Developmental Screening and Surveillance* uses an instrument called PEDS (Parental Evaluation of Development Status). This has been trialled in Australia since 2003<sup>2</sup>. In the present context, it is used mainly as a surveillance rather than a screening tool (and therefore presumably for collecting population rather than individual data). However the *Handbook* includes the statement: "If the PEDS raises concerns about the child's development then a more formal developmental assessment needs to take place before clinical conclusions can be made. No recommendation

---

1 ASHA (1997) *Guidelines for Audiologic Screening*. PLACE: American Speech-Language-Hearing Association. URL: <http://www.asha.org/policy>

2 Australian Institute of Family Studies (2003) *The Platform Strategy*, City of Moonee Valley, Victoria, *Stronger Families Learning Exchange Bulletin*, 5, 47-48.

is made about the type of secondary level assessment that should be performed, or by whom it should be undertaken. However several services are using the *Ages and Stages Questionnaire (ASQ)*.”

- (vi) *Growth Measuring and Monitoring* (10 pages in Handbook). This includes height and weight monitoring, and reporting of B.M.I. based on these measures. (B.M.I.= “Body Mass Index”, computed as body mass divided by the square of height). Referrals are recommended: (a) If a child is less than the 3<sup>rd</sup> percentile for height and weight using WHO growth standard. (b) If a child has a B.M.I. of 21 or more; “Refer to GP for on-going weight monitoring and management of complications of obesity”; (c) If a children is overweight but has a B.M.I. less than 21: “Do not refer overweight children whose B.M.I. is less than 21, but give parent information on healthy eating and healthy activity. Referral not recommended because there is limited evidence.” However, it is well known that Maori and other Polynesians have a distinctive body form<sup>3</sup>. B.M.I. cut-offs based on average Caucasian body form may incorrectly identify obesity in Maori or Pasifika people (where it is falsely assumed that it represents body fat rather than muscle). Of many studies on this (including one in children and adolescents) all but one recommend different cut-offs for Maori/Pasifika compared with other groups. There is widespread concern about type II diabetes and related metabolic problems in Maori/Pasifika. However, many studies also show either that the association between obesity and type II diabetes and/or metabolic problems disappears when an appropriate cut-off is used, or that risk factors other than body form account for the elevated rate of these problems in Maori/Pasifika. Use of B.M.I. in the *B4 School Check* makes no mention of this: B.M.I. is taken to have the same significance for all ethnic groups.

---

3 WHO guidelines take no account of the distinctive body form of Polynesians and Maori. This has been a matter of controversy. see: Swinburn,BA, Craig,PL, Daniel,R, Dent,DP, Strauss,BJ (1996) Body composition difference between Polynesians and Caucasians assessed by bioelectric impedance.. *International Journal of Obesity and Related Metabolic Disorders*, 20, 889-894; Rush, EC, Plank,LD, Lauulu,MS, Robinson,SM (1997) Prediction of percentage body fat from anthropometric measurements: comparison of New Zealand European and Polynesian young women. *American Journal of Clinical Nutrition*, 66, 2-7; Salesa,JS, Bell,AC, Swinburn,BA (1997) Body size of New Zealand Pacific Islands children and teenagers. *New Zealand Medical Journal*, 110, 227-229; Rush,EC, Plank,LD, Robinson,SM (1998) Estimation of body fat in Caucasian and Polynesian women from anthropometric measurements. *Applied Radiation & Isotopes*, 49, 749-750; Swinburn,BA, Ley,SJ, Carmichael,HE and Plank,LD (1999) Body size and composition in Polynesians. *International Journal of Obesity & Related Metabolic Disorders*. 23, 1178-1183; Craig,P, Samaras,k, Freund,J, Culton,N, Halavatau,V, Campbell,L. (2003) BMI inaccurately reflects total body and abdominal fat in Tongans. *Acta Diabetologica*, 40, suppl1, S282-285; Rush,EC, Freitas,I, Plank,LD (2009) Body size, body composition and fat distribution: comparative analysis of European, Maori, Pacific Island and Asian Indian adults. *British Journal of Nutrition*, 102,632-641; Rush,EC, Crook,N, Simmons,D. (2009) Optimal waist cut-point for screening for dysglycemia and metabolic risk: evidence from a Maori cohort. *British Journal of Nutrition*, 102, 786-791; Taylor,RW, Brooking,L, Williams,SM, Manning, PJ, Sutherland, WH, Coppell,KJ, Tipene-Leach,D, Dale, KS, McAuley,KA, Mann,JI (2010) Body mass index and waist circumference cutoffs to define obesity in indigenous New Zealanders. *American Journal of Clinical Nutrition*, 92, 390-397.



## **B: What has Driven the Mental Health Components of the B4 School Check?**

*Overt objectives.* In New Zealand, child health is guided by an evolving set of documents entitled *Well Child/Tamariki Ora Framework*. In February 2006 the government started a review of this framework, and in August 2007, the Ministry of Health published the resulting document (*Supporting Evidence for Changes to the Content of the Well Child/Tamariki Ora Framework*). Areas needing more emphasis in health promotion were (a) maternal postnatal depression; (b) child maltreatment; (c) obesity, nutrition and breastfeeding; (d) dental care and oral health; (e) child mental health and attachment; (f) developmental delay and behavioural problems. Mention of child maltreatment, child mental health and attachment, and behavioural problems no doubt arises from increased awareness in New Zealand in recent years of large numbers of severely dysfunctional families, along with sky-rocketing levels of family violence, some of it very severe, with all its consequent impact on children of those families. The 2007 review includes the following statement:

“Child development is a powerful determinant of health in adult life. Developmental surveillance is a shared activity between parents and health professionals, and uses both parties’ knowledge of the child to monitor the child’s ongoing development. The Ministry is proposing to use a widely accepted validated screening tool, Parental Evaluation of Developmental Status (PEDS), both to detect developmental and behavioural problems and as a means of eliciting and responding to parental concerns.”

Recommendations included new or revised screening systems for vision, hearing (in view of concern about “glue ear” in children), oral health, and introduction of a new questionnaire to identify behavioural or developmental concerns. From these concerns, the *B4 School Check* was devised, envisaged to take place at ages between four to five-and-a-half years (preferably four to four-and-a-half years). This was to include (in addition to general health assessment, oral health assessment, hearing and vision screening, and measurement of height and weight), recording of body mass index (B.M.I.) as a population-level indicator<sup>4</sup>, and screening to detect behavioural and developmental problems through use of the *Strengths and Difficulties Questionnaire* (SDQ) and the PEDS.

*Health Screening: Guidelines and Dilemmas:* Health screening programmes often arouse controversy and are (and should be) matters for public debate. In view of this, as background to the *B4 School Check*, guidelines of the National Health Committee guidelines for current or future health screening programmes<sup>5</sup> should be mentioned. Screening programmes should meet eight criteria: (a) The condition is a

---

4 Clearly what was originally intended as an aggregate (population level) indicator, became one for individual assessment and referral by the time *B4 School* was deployed, in view of the referral guidelines contained in the *Practitioner’s Handbook*.

5 National Health Committee, Wellington (2003) *Screening to Improve Health in New Zealand. Criteria to assess screening programmes*. National Advisory Committee on Health and Disability, April 2003.

suitable candidate for screening; (b) There is a suitable test for the condition; (c) There is an effective and accessible treatment or intervention for the condition identified through early intervention; (d) High-quality evidence, ideally from randomised controlled trials, shows that a screening programme is effective in reducing morbidity or mortality; (e) The potential benefit from the screening programme outweighs the potential physical and psychological harm caused by the test, diagnostic procedures and treatment; (f) The health care system is capable of supporting all necessary elements of the screening pathways, including diagnosis, follow-up and programme evaluation; (g) Social and ethical issues are considered; (h) Cost-benefit issues are considered. These guidelines acknowledge that there is usually a fine balance between benefits and harms in proposed screening programs and emphasise the need for fully informed choice by those invited to take part in a screening program.

There is no reason to oppose health screening in principle, and many reasons to support it when properly developed. However some questions are already raised: (i) The guidelines for referral in the *Practitioner's Handbook* appear to be more precise and thorough for hearing, oral health, and perhaps vision than for the behavioural and developmental screening components and for the SDQ. (ii) One may question whether the PEDS system meets criteria (c) and (d). (iii) One may also ask whether use of B.M.I. in the *Growth Measurement and Monitoring* section of *B4 School* meets criterion (b). (iv) A distinction might be made between assessment for conditions for which benefits (initially at least) are felt at an individual level, and those (such as immune status for infectious disease), where population-level screening contributes to reducing risk of an epidemic. In the latter case, where “herd immunity” can protect a whole population, some would argue that usual procedures for signed informed consent should be waived or modified<sup>6</sup>.

In the section of *B4 School* on behaviour problems, more specific problems arise. Some community voices have expressed concern about *B4 School*, and particularly use of the Strengths and Difficulties Questionnaire, suggesting that it may be driven by a hidden agenda, especially in screening for mental health and behavioural issues in pre-schoolers<sup>7</sup>. This serious charge needs to be examined carefully. In addition, it is understood that health screening can be carried out at various levels – universal, targeted at groups at special risk, or screening closely linked to existing clinical services. A review of this<sup>8</sup> discussed merits and disadvantages of each approach,

---

6 Kissam,S, Gifford,DR, Patry,G, Bratzler,DW. (2004) Is signed consent for influenza or pneumococcal polysaccharide vaccination required? *Archives of Internal Medicine*, 164, 13-16; Malmqvist,E, Helgesson,G, Lehtinen,J, Natunen,K, Lehtinen, M. (2011). The ethics of implementing human papilloma virus vaccination in developed countries. *Medicine, Health Care Philosophy*, 14, 19-27.

7 Maria Bradshaw: *Universal Psychiatric Screening for NZ Pre-Schoolers*. (see: Mad in America: Science, Psychiatry and Community)

<http://www.madinamerica.com/2012/03/universal-psychiatric-screening-for-nz-pre-schoolers/>.

I decided not to read this and similar reports until I had completed my own report. I wanted my own report, as far as possible, to be independent of any other critique of *B4 School*.

8 Offord,DR, Kraemer,HC, Kazdin,AE, Jensen,PS, Harrington, R. (1998) Lowering the Burden of suffering from child psychiatric disorder: Trade-offs among clinical, targeted and universal interventions. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 686-694.

and concluded that a mixture of the three is usually best. One may ask whether this advice is followed in deployment of the SDQ, in *B4 School*. Discussion of this is left until later, after a detailed review of procedures used in the *B4 School Check*.

## **C: How the Objectives of *B4 School Check* are presented.**

**[I] History of Screening for Behavioural/Emotional Difficulties in Pre-Schoolers and Children.** Antisocial behaviour, especially amongst adolescents has long been a matter of concern for parents, the wider public, and those who develop social policies. Such behaviour can be very costly. Before 1980, little thought was given to the possibility of detecting such behaviour early, and putting in place interventions to reduce its occurrence or severity. However, around that year the idea was discussed<sup>9</sup> that psychological screening in young people might identify those most at risk, opening up the possibility of taking preventive measures. Since then, further research revealed risk and protective factors for such behaviour, and it became clear that these problems were often rooted in a child's early development. In the last 15 years these insights, have led to a number of attempts not only to detect those at risk at an early stage, but also to implement preventive interventions, either by helping parents develop better parenting styles, or by guiding children themselves towards better styles of social interaction. Most such programs to date are research studies, rather than established programs of intervention. As with any public health measure, there is a long journey between showing that effective intervention is possible in principle in controlled research studies, and finding ways to implement the principles on a population-wide scale. Many ideas for intervention which seem good in theory, fail in practice, for a wide variety of reasons.

### **[II] *Apparent Objectives of Behavioural/Developmental Component of B4 School Go Beyond Those Advised in its Supporting Documents.***

*B4 School* is presented in different ways to different constituencies. The paper for the education sector is very different and less substantial than the *Handbook for Practitioners*, intended to guide those who implement or coordinate assessments. The following section is based on pp. 27-41 of the Handbook. Behavioural and developmental screening is clearly seen as an important part of the whole program of screening, in view of the length devoted to it in the Handbook.

On page 27 we read: "There is evidence that early identification and intervention improves developmental and social outcomes for the child and family/whanau and the earlier the intervention the better". The reference here is "Centre for Community Child Health, Royal Children's hospital, Melbourne (2002)."<sup>10</sup> However, there are two causes for concern in citing this report:

---

9 Nilsson, C, Sundelin, C, Vuille JC (1977) General health screening of four-year-olds in a Swedish country. V. A Strategy for improving the effectiveness and the cost efficiency of the psychological screening program. *Acta Paediatrica Scandinavica*, 66, 289-295; Brink, SG, Nader, PR (1984) Comprehensive health screening in elementary schools: an outcome evaluation. *Journal of School Health*, 54, 75-78.

10 *Child health surveillance and screening: a critical review of the evidence*. Australia: National Health and Medical Research Council. <http://www.nhmrc.gov.au/publications/synopses/ch42syn.htm>

- (a) In the synopsis of this Australian report (based on information available in 2000) it states: “The early detection of health and other problems in children is a worthy goal. At first glance the benefits appear to be self-evident. It is not until one begins to systematically review the evidence for screening and surveillance, topic by topic, that the complexity of this endeavour becomes apparent. There is surprisingly little evidence for the effectiveness of screening programs in many domains. . . . There are scant data about cost effectiveness. There are major issues of program quality, monitoring of compliance with referrals for assessment, and whether facilities exist in many communities for assessment and follow up. There are concerns that much attention is paid to the test or procedure itself, and little to the elements of a community-wide program. In some cases there is little evidence that therapy alters outcomes.” These comments were made on the basis of evidence available in the year 2000. Evidence available now may be better, and more recent evidence is discussed in a later section.
- (b) Although the chapter of the *Practitioner’s Handbook* is entitled “Identifying developmental and behavioural problems” and half of it is about behavioural difficulties (and use of the SDQ), the Australian document has a chapter on developmental problems, but not one on behavioural problems.

**[III] *Disconnection Between Different Parts of Ministry of Health’s Agenda for Child and Youth Mental Health.*** As described in the section [B] above, *B4 School* is part of the Ministry of Health’s agenda for children and youth health. While most of its methods are well tried, some are still under evaluation (see below). However, another recent document from this Ministry, published in December 2012, addresses the mental health/addiction sector across age groups (*Rising to the Challenge: The Mental Health and Addiction Service Development Plan, 2012-2017*). Chapter 5 of this document is entitled “Delivering increased access for infants, children and youth while building resilience and averting future adverse outcomes.” Clearly it is concerned with preventive as well as other aspects of child mental health, and one would expect it to be coordinated with, or at least to refer to, the behavioural section of *B4 School*. However the latter program is not mentioned. Chapter 5 includes the following statement: “. . . there is good evidence that intervening effectively and early with infants, children and young people works, can avert more serious issues in the future and is highly cost-effective”. The reference here is to a document from the Office of the Prime Minister’s Science Advisory Committee, 2011. The relevant document<sup>11</sup>, from the Chief Science Advisor’s Office, is mainly about problems starting in early adolescence, but in its Chapter 4 on Childhood Conduct Problems (p.61) the prevalence of this problem in the 3-17 year age group is mentioned. The only mention of *B4 School* (p. 184-185) is this: “Population norms for children across all estimable ages of risk assessment (4-16 years) do not exist in New Zealand. Data are being collected as part of the *B4 School Check*, but only at this young age.” Thus, different parts of the Ministry of Health seem little aware of each other’s agendas.

---

11 Office of the Prime Minister’s Science Advisory Committee, Auckland. (2011) “*Improving the Transition. Reducing Social and Psychological Morbidity During Adolescence: A Report from the Prime Minister’s Chief Science Advisor. May 2011 (accessed at: www.pmcsa.org.nz)*”

**[IV] Disconnection Between Aims and Methods of Behavioural Testing in B4 School, and Supporting Documents.** On page 27 of the *Practitioner's Handbook*, we read: "Management of behavioural problems may be effective in 75-80% of preschool children with antisocial behaviour. Once the child is aged five to seven, management is effective in 65-70%, but once the child is aged 8-12 years this decreases to 45-50%. It is very difficult to change antisocial behaviour in adolescents." The reference cited here is a document sometimes referred to as the Church Report<sup>12</sup>: Several points arise from this citation:-

- (a) It is clear in the Church Report that the area of concern is anti-social behaviour, which, as child psychiatric diagnoses (according to DSM-III or DSM-IV) equates to Oppositional Defiant Disorder, Conduct Disorder, and Attention Deficit Hyperactivity Disorder (whose definitions are discussed). None of those diagnoses are mentioned in the *Practitioner's Handbook for B4 School*. It appears to avoid such terms, despite their being freely used in the document from which *B4 School* arose. Nevertheless, concern about such problems seems to be a prominent motive for using the SDQ in *B4 School*.
- (b) The Church Report never mentions the SDQ - the instrument to be used in *B4 School* for assessing behavioural problems<sup>13</sup> - although this instrument had been published several years before that report was finalised<sup>14</sup>. Thus, there is a "disconnect" between the Church report and *B4 School*, which belies the way the former is cited to support the latter.
- (c) The authors of *B4 School* seem equivocal about whether the behavioural assessment is to be seen in a medical, educational or other paradigm, and whether the aim is to screen for statistical risk, or to assess individuals with a view to referral on the basis of risk. In the Church Report, section 3 discusses the relative merits of medical, developmental and educational methods of assessments, and section 7 advocates "functional assessment", in preference to medical or educational assessment. Nevertheless the medical term "diagnosis" is often used, not least in the title of the report (despite the report

---

12 Church,R. (2003) *The definition, diagnosis and treatment of children and youth with severe behaviour difficulties: A review of research*. Wellington, Ministry of Education.

13 This is not the only occasion on which there has been omission of reference to the behavioural screening, while retaining reference to the developmental screening part of *B4 School*. In response to a question in the New Zealand parliament by Tim McIndoe (National MP, Hamilton West) about use of *B4 School* ("What progress has the government made toward ensuring all eligible New Zealand four-year-olds get a *B4 School Health Check*?"), the Minister of Health Mr Tony Ryall responded as follows: "*B4 School* checks are a comprehensive free health and development check for 4-year-olds to make sure they are well set up for school. I can advise the House that in the past 2½ years children have received over 100,000 *B4 School* checks. About one fifth of these kids have been identified with eyesight, hearing or developmental problems and have been referred to appropriate services. In making this program work better, we are not only making kids health better but improving their education and their ability to interact in the community. I contrast the 100,000 in the last 2½ years with the 3,000 achieved by the previous government before it left office. *Hansard* 13.09.2011 (p 21184). (The fact that the *B4 School* check, although planned by the former government, was not deployed until shortly after it left office seemed to have escaped the Minister.)

14 Goodman,R, Renfrew,D and Mullick,M (2000b) Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child & Adolescent Psychiatry*, 9, 129-134.

having been commissioned by the Ministry of Education). Perhaps, it could be argued, medical terminology is used as “common language” rather than in a more technical sense. Nevertheless, the word “diagnosis” *does* imply individual assessment, rather than the looser concept of statistical assessment of risk. The implication is clear on p. 50 of the Church report (“Summary, conclusions, and unresolved issues”) where the first summary point reads: “In order to prevent antisocial children growing up to become antisocial adults it is desirable that such children be identified as early as possible and as soon as the signs of antisocial development begin to appear. Achievement of this goal will require diagnostic screening procedures which are more accurate than those currently in use.”

The objective of the behaviour/development part of *B4 School* is stated clearly to be for screening, not individual “diagnosis” in a strict medical sense. So, we read (*Practitioner’s Handbook*, p. 34): “The SDQ is a screening tool; it is not diagnostic. Therefore it is only an initial indication of whether a child and their family/whanau may need support and/or intervention”. Again, in Appendix 2 (p. 56), we read: “The difference between the checks conducted as part of the *B4 School Check* and referral to diagnostic services (for hearing, vision, general health (including oral health), behavioural or developmental concerns) includes an explanation that the various screens conducted are screens only and as such the inherent limitations in conducting such screens.” However, in a shorter document for the education sector, this distinction is not explained, and we read: “If concerns are identified during the check, the child and their parents will be offered support to improve the child’s health and well-being. This could include referral to specialist services”.

Despite these disclaimers, the stated aims of *B4 School* do not correspond exactly to the intended aims of the SDQ. The SDQ *is* intended to give 4 “broad-band diagnoses”. Its authors, from Institute of Psychiatry in London, were clearly thinking in a medical context. One of their early papers on the SDQ is entitled: “Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka.” They write, with respect to ADHD: “This minority is particularly likely to benefit from the involvement of a child psychiatrist or paediatrician trained in the recognition and pharmacotherapy of hyperactivity disorders”<sup>15</sup>. There is another “disconnect” here between the aims of the SDQ, and those of *B4 School*.

[V] ***Which Aspects of Child Mental Health Does B4 School Screen For?*** The four broad-band diagnoses which the SDQ attempts to identify are rated in four scales: Emotional Symptoms, Conduct Problems, Hyperactivity, and Peer Problems. There is also a “Prosocial Scale”, a Total Difficulties Score, and an Impact supplement. This is a rather narrow range of child mental health problems. For instance, on the Barnardos website<sup>16</sup>, early signs of mental illness are listed as

- difficulty getting along with children their own age
- extreme fear, and/or anxiety

---

15 Goodman,R, Renfrew,D and Mullick,M (2000b) Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child & Adolescent Psychiatry*, 9, 129-134.

16 <http://www.barnardos.org.nz/Family%20Advice/Child%20mental%20health/Child%20mental%20health%20signs>

- lack of energy or motivation
- irritability, restlessness, fidgeting and trouble concentrating
- very odd or unusual behaviour
- obsessive or ritualistic behaviour
- changes in school performance
- changes in sleeping patterns and/or eating patterns
- spending less time with, or avoiding, friends
- excessive aggression, breaking the rules, truancy, theft, vandalism
- prolonged negative mood, perhaps crying a lot and feeling hopeless or worthless
- substance abuse or other self-destructive behaviour
- self-harm or suicide attempt.

Population prevalence of various child psychiatric disorders is discussed below, when the merits of the SDQ are considered in more detail (section E, [IV]).

**[VI] *Roots of Anti-Social Behaviour in Children: An Individual or a Societal Problem?*** Assuming that the objective of the behavioural screening component of *B4 School*, is assessment and intervention at an individual level, a fundamental issue of political philosophy arises: To what extent do problems of antisocial behaviour in youngsters arise in *individuals or their families*, rather than at a *societal level* related to social mores and social/economic conditions, poverty, and policies or styles of particular environments created for children (including those created in schools)? In asking this question it is not intended to question the merits of intervention at an individual level. Good or bad behavioural patterns are often established at an early age, and once they start to form, they may be steadily reinforced both from within a developing child, and by external influences. This being so, intervention at an early age, at an individual level, *is* likely to be part of any comprehensive solution. But is it the only level at which the issue can be addressed? In today's individualistic world, governments and their agencies, of whatever political stripe, may be persuaded by Margaret Thatcher's one-liner: "There is no such thing as society, only the individual"<sup>17</sup>. However, in other areas of behavioural medicine, such as suicide prevention, some of the best programs operate at the level of entire communities<sup>18</sup> (with strategies at the individual level still included). This issue also has an international dimension, since several of the early studies intended to validate the SDQ were conducted in poor countries (Pakistan<sup>19</sup>, Bangladesh<sup>20</sup>, Democratic

---

17 . . . a line, which, curiously enough, probably originated (in German) with Carl Gustav Jung (see Ellenberger, H, 1970, *Discovery of the Unconscious*, Basic Books, p. 682.) who, of course, meant something entirely different.

18 See: Hatim,O. (2005) A model program for youth suicide prevention. *International Journal of Adolescent Medicine & Health*, 17, 275-278; for current websites, see: stopyouthsuicide.com; <http://www.spreadthelovesavealife.org/> <http://uknow.uky.edu/content/spread-love-save-life>

19 Samad,L, Hollis,C, Prince,M, Goodman,R. (2005) Child and adolescent psychopathology in a developing country: testing the validity of the Strengths and Difficulties Questionnaire (Urdu version). *International Journal of Methods in Psychiatric Research*, 14, 158-166.

20 Mullick,MSI and Goodman,R. (2001) Questionnaire screening for mental health problems in Bangladeshi children: a preliminary study. *Social Psychiatry and Psychiatric Epidemiology*, 36, 94-99.

Republic of the Congo<sup>21</sup>). In such countries, child mental health problems may best be seen as part of much larger issue of poverty and under-development. One may then ask whether child mental health issues should be addressed by assessing and treating individual children or by political change on a much bigger scale. The same applies to some extent in New Zealand.

In the *Practitioner's Handbook for B4 School*, this is admitted, where, in relation to developmental delay we read (p. 35): “Children in poverty are particularly vulnerable. Low socioeconomic status during childhood interferes with a child’s cognitive and behavioural development and is a modifiable risk factor for lack of readiness for school.<sup>22</sup> We also read (p. 36): “Factors such as poverty, low parental education levels and high levels of parental stress are associated with a greater risk of language delay. Despite this, few studies have adequately documented the incidence of language delays in the presence of such factors.” Of direct relevance to the use of the SDQ within *B4 School Check*, a study from Germany, using the German version of the SDQ<sup>23</sup> in 3-17 y-olds, found that scores on all subscales and total score were markedly worse for children and adolescents of low compared to high socio-economic status.

Despite these comments, and admissions that the evidence is inadequate, the emphasis seems to be entirely on assessment (and intervention) at the individual level. This may be fine for some problems identified in the *B4 School Check*, but it is a concern that this emphasis prevails for behavioural tests, where broader society-wide influences are likely to be in play. One could take the problems of individual youngsters back to their parents, then to *their* upbringing, and ultimately to the social environment of whole countries (and to international trends) developing over the last generation. Social problems may be seen to derive from the entire model of society promoted by governments and their agencies of all shades, in today’s world. At least some of the remedies should work at the societal level.

In view of these comments one seeks the broader context of New Zealand social policy. A document recently released (“White Paper for Vulnerable Children”<sup>24</sup>) is concerned exclusively with data collection, identification of those at risk, and with intervention and prevention at an individual level. In her introductory statement the Minister writes: “Though I acknowledge the pressure that financial hardship puts on families, that is never an excuse to abuse children. Most people in such circumstances do not abuse their children, and I cannot tolerate it being used as

---

21 Kashala,E, Elgen,I, Sommerfelt, K, Tylleskar, T. (2005) Teacher ratings of mental health among school children in Kinshasa, Democratic Republic of Congo. *European Child & Adolescent Psychiatry*, 14, 208-215.

22 Hertzman 2001, cited in Anderson et al, [2002] Community interventions to promote healthy social environments: early childhood development and family housing; a report on recommendation of the Task force on Community Preventive Services. *MMWR Recommendations and Reports*, 51, (RR01) 1-8.

23 Hölling,H, Kurth,B-M, Rothenberger,A, Becker,A, Schlack,R. (2008) Assessing psychopathological problems of children and adolescents from 3 to 17 years in a nationwide representative sample: results of the German health interview and examination survey for children and adolescents. *European Child & Adolescent Psychiatry*, 17, suppl 1, 34-41.

24 Paula Bennett, (2012) *The White Paper for Vulnerable Children*. Ministry for Social Development, Wellington.



justification for those who do.” This comment is somewhat naïve on the social roots of the problem behaviours, with breakdown of communal bonds which help people (including future parents) to become strong in themselves, with an in-built sense of communal and family ties, morality, and responsibility to a wider community. This is fostered by a free-for-all competitive environment prevailing in so many aspects of today’s societies. The Minister seems to be washing her hands of all this.

Two recent reviews place emphasis on social rather than individual factors, and on life events often related to these social factors, rather than generic childhood predictors, as determinants of adolescent delinquency, and later criminality. In one of these<sup>25</sup>, based on experience in very deprived areas of Teesside, UK, we read: “Living in poor communities heightens and concentrates risk, while resources to cope with life events such as bereavement are diminished” . . . “Much of what happened in the lives of our informants could not have been predicted from earlier experiences.” “Criminal career research and its operationalization in risk assessment devices have not taken sufficient account of the role of accelerated social and economic change in engendering and concentrating risk factors in destabilized neighbourhoods among their inhabitants. Neither do they take account of unpredictable life events. In isolating individual risk factors from their context in biography, place and social structure, such devices offer ways of managing offenders rather than addressing the causes and cessation of individual offending.”

In similar vein, the second review<sup>26</sup> offers the following commentary: “Comprehensive models of conduct disorder point toward distal factors such as neighbourhood-level violence, cultural endorsement of aggression, culture of peer deviance, discrimination against specific groups, and socioeconomic class divides between high-risk families and mainstream institutions of education and juvenile justice. Individual-level interventions may take into account these factors, but cannot easily alter these factors. Clinicians ‘work around’ or ‘work with’ community risk factors; they almost never work to change these factors. Going to scale with individual-level interventions may ignore cultural and community causes, leading to the perpetual replication of new cases with little net impact on community rates of problems.”

**[VII] A Hidden Agenda?** “Early-intervention” is a currently-fashionable “buzz-word”. However, in the case of screening with the SDQ, one should ask: Early intervention to prevent exactly *what*? The *B4 School* documentation avoids direct mention of medical/psychiatric terminology for behavioural or emotional problems in pre-schoolers. However, by comparing wording in sections of the *Practitioner’s Handbook* with that in other contemporaneous documents, its provenance becomes clear, and this question can be answered. In the *Practitioner’s Handbook*, p. 27, we read: “Management of behavioural problems may be effective in 75-80% of preschool children with antisocial behaviour. Once the child is aged five to seven, management is effective in 65-70%, but once the child is aged 8-12 years this decreases to 45-50%. It is very difficult to change antisocial behaviour in

---

25 Webster,C. (2006) Predicting criminality? Risk factors, neighbourhood influence and desistance. *Youth Justice*, 6, 7-22.

26 Dodge, KA (2009) Community intervention and public policy in the prevention of antisocial behavior. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 50, 194-200.

adolescents.” Compare this with the following sections in the document entitled *Inter-Agency Plan for Conduct Disorder/Severe Antisocial Behaviour*<sup>27,28</sup>, (published September, 2007, 13 months before the *Practitioner’s Handbook*):-

p.5: “Conduct disorder/severe antisocial behaviour, particularly in younger children, is also one of the strongest predictors of poor long-term outcomes into adulthood, including criminal offending, substance abuse, and mental health problems.”

p. 30: “The onset of conduct disorder/severe antisocial behaviour in early childhood is one of the strongest predictors of juvenile delinquency and behavioural difficulties into adulthood.”

p. 31: “Interventions are significantly more effective for younger children than their older counterparts. Prior to school entry, it is estimated that there is a 75-80% chance of halting exiting antisocial behaviour in young children and increasing pro-social behaviour [citing Church 2003] Programme effectiveness drops sharply with age.”

p. 43 (recommendation): “Implement systematic screening and assessment processes for all 3-7 year-olds within the education sector” (part of Phase Two, 2009-2010. This can only refer to *B4 School*, although not named as such).

A related document is entitled *Drivers of Crime: Background Information*<sup>29</sup>, published in November 2009. In this document (p.8) we read “Serious behavioural problems in early childhood are one of the strongest predictors of adverse adult outcomes, particularly offending. Between 5 and 10 per cent of children are estimated to have severe behaviour problems. The incidence of serious behaviour problems is estimated to be 10 to 15 percent among Maori children. Similarly, the *Inter-Agency Plan* states: “New Zealand research suggests that Maori and Pacific males are more likely to have difficulties than non-Maori, though to a large extent this is likely to be due to economic disadvantage [citing Fergusson et al, 2004]<sup>30</sup>”

As detailed later in this report, use of the SDQ just for the youngest age-group, is unprecedented, and, in this age group, not properly validated. Readers may suspect

---

27 Church,J, Langley,J, Liberty,K, Poulton,R, Ronan,K, Werry, JS. (2007) *Inter-Agency Plan for Conduct Disorder/Severe Antisocial Behaviour, 2007-2012*. Ministry of Social Development, Wellington.

28 A key paper cited as showing that conduct disorder in children is one of the strongest predictors of crime in adulthood is Fergusson DM. Horwood LJ. Ridder EM. (2005) Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood. *Journal of Child Psychology & Psychiatry & Allied Disciplines*. 46, 837-49. The content of this paper is discussed in section E,VII,(iv), and Table IV, below.

29 This report from the Ministries of Justice and Maori affairs, entitled “Addressing the Drivers of Crime: Background Information” developed from a meeting in April, 2009, hosted by the Ministries of Justice and Maori Affairs, was attended by 100 persons from various backgrounds with common interest in preventing crime and victimization. On 2nd November, 2009, the cabinet agreed that “Addressing the Drivers of Crime” become a whole-of-government priority.

30 The *Interagency Plan*’s authors are Ministries of Education, Health, Justice, Social Development (including Child Youth and Family, and the High and Complex Needs Unit), and the following individuals: Dr. John Church, Prof David Fergusson, Dr John Langley, Dr Kathleen Liberty, Prof Richie Poulton, Prof Kevin Ronan, Prof John Scott Werry. It is not clear what input (if any) there was from genuine Maori health professionals.

that it was included in haste and prematurely, its intentions concealed, as a surreptitious part of the “Drivers of Crime” agenda.

## **D: What were the intended ways of implementing the SDQ, and of using information derived from it?**

[I] *By Whom, and How is Rating via SDQ conducted?* Assessment of each child is by their parent (“SDQ-P”) or their teacher (“SDQ-T”: usually from a pre-school teacher). Whether parent or teacher provides information, a practitioner (usually of nursing background) is present to supervise and guide the process, but parents, if they wish, can fill in the form unaided. In so far as practitioners with a nursing background supervise the process, it is not clear that they would have experience or specialty training in child health (as would have been the case in the days of the School New Entrant Check). Research data make it clear that the SDQ is more reliable when both teacher and parent provide reports, and the *Practitioner’s Handbook* (p.34) states: “The SDQ is considerably more sensitive when both SDQ-P and SDQ-T have been completed than when only one has been completed.” Authors of the SDQ also provide an algorithm for combining results of the different reports to give a more reliable assessment<sup>31</sup>. As already mentioned in the document from August 2007, the Ministry of Health prior to launch of the *B4 School* system advised that “developmental surveillance is a shared activity between parents and health professionals, and uses both parties’ knowledge of the child to monitor the child’s ongoing development.” However, there is no evidence that this algorithm has been used in New Zealand, and statistics (detail in section G) give numbers screened separately on the basis of teacher and parent reports, with no evidence of combination of scores. The practice of asking parents to assess their own children is said in the *Practitioner’s Handbook* to have two advantages: “It involves family/whanau in the identification and management processes and is cheaper than involving health professionals”<sup>32</sup>. In contrast, a recent study in the USA intended to validate the SDQ<sup>33</sup> collected data through computer-assisted personal interviews conducted by trained U.S. Census Bureau interviewers. Another view from the USA is of interest: Allen Frances, a strong public critic of DSM-V (published, May 2013) and of what he sees as a trend to “diagnostic inflation” is concerned about use of lay interviewers to keep costs down. He suggests that there is then no filter for “clinical significance” of distress or impairment, which inflates the rates of many disorders. While his concern is about epidemiological studies, it may apply to mental health screening programs, leading to an excess of false positives (section E[VI]).

---

31 Goodman,R, Renfrew,D and Mullick,M (2000b) Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child & Adolescent Psychiatry*, 9, 129-134.

32 Tsiantis,J, Smith,M, Dragoness,T, Cox,A. (2000) Early mental health promotion in children through primary health care services: A Multi-Centre Implementation. *International Journal of Mental Health Promotion*, 2 (3), 5-17.

33 Bourdon,KH, Goodman,R, Rae,DS, Simpson,G, Koretz,DS. (2005) The Strengths and Difficulties Questionnaire: US Normative data and psychometric properties. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 557-564.

[II] *Advice to Practitioners to Use of Language to Hide the Real Objectives of the SDQ: Issue of Truly Informed Consent.* Another problem in implementing the SDQ is the recommendations on how language is to be used. In the *Practitioner's Handbook* (p.33) we read that the score for each of the four subscales is to be reported, plus the warning: "IMPORTANT: Do not use the words 'borderline' or 'abnormal' when communicating the results back to the parent. Use the terms 'concerning' to acknowledge parent's concerns". Clearly the authors of *B4 School* are sensitive to possible labelling or branding someone as abnormal in the area of mental health, although not in other parts of B4 School, where the word "abnormal" is not avoided. It is easy to see through this ploy: In other parts of the *Practitioner's Handbook*, (e.g. p.34), it is admitted that the word "concerning" equates to the word "abnormal" in previous documents. Moreover, in documents easily available online<sup>34</sup>, the terms "Abnormal" and "Borderline" are used freely as official categories in reporting statistics on use of the SDQ. Clearly, in terms of the Guidelines for Health Screening, such advice to practitioners on how they interact with parents and/or teachers falls short of "fully informed choice to participate".

On this, an important precedent comes from the HIV/AIDS area: When an antibody-based test for diagnosing HIV-status was developed, the issue was hotly debated in San Francisco in the mid-1980s, in relation to population-wide screening<sup>35</sup>. The policy which emerged, to become a model for every other US city, and eventually "sold" to Federal Health Authorities by a wily fundamentalist pastor was as follows: (a) The test should be voluntary; (b) There should be written consent; (c) Legislation was passed to make it a criminal offense to break confidentiality on test results. Without such safeguards, it was judged that it would not be acceptable to the public. The politics relating to adoption of this policy was extremely intense<sup>36</sup>, and it was more open and transparent than has been any debate about questionable parts of *B4 School* in the New Zealand context. In my view, such openness, and the trust in the public it implies, is the only way to introduce health screening for sensitive topics such as risk of mental disorders.

In the New Zealand context two points can be made here on the use of the SDQ: (i) There is ambivalence between the supposed voluntariness of giving consent, and the intention of administrators and politicians driving the program, evident from many sources, that screening should be near-universal. Indeed, DHBs have been given performance targets to fulfil, specifying the desired percentage of the age group to have been screened by a given date<sup>37</sup>. Pressure may be applied by practice nurses on parents<sup>38</sup>, or conveyed "from above" to practitioners. For instance, a news

---

34 *B4SC Quality Letter* (August 2012), available at:

<http://www.scdhb.health.nz/uploads/Primary%20and%20Community/CPHAC%20Dec%207%204%20B4%20School%20Checks%20Newsletter.pdf>

35 Randy Shilts (1987) *And the Band Played on: Politics, People and the AIDS Epidemic*. St.Martin's Press.

36 See Siplon, PD (2002) *AIDS and the Policy Struggle in the United States*. Georgetown University Press, Washington DC.

37 Initially the target was 80%, which recently shifted to 90%

(see: <http://www.beehive.govt.nz/release/record-number-children-get-b4-school-check>)

38 [http://www.beyondconformity.org.nz/hilarys-desk/a\\_reader\\_writes\\_in\\_b4\\_school\\_check](http://www.beyondconformity.org.nz/hilarys-desk/a_reader_writes_in_b4_school_check)

report<sup>39</sup> expressed concern about pre-schoolers who had missed the audiometry part of *B4 School*. Recent reports<sup>40</sup> claim that practitioners were systematically “cutting corners” in the Hearing Screening test in some DHBs, by testing only one of a child’s ears, or using results from a health professional’s own ear instead of those from the child. That consent is given in truly voluntary fashion by fully informed parents is called into question by the fact that The Ministry of Health has entered into partnership with producers of a television cartoon program (*Wots Wot*) aimed at pre-schoolers to promote the *B4 School* checks<sup>41</sup>. Such promotional campaigns sit uneasily with the notion of truly informed consent. Such ambivalence may be intrinsic to the thinking behind *B4 School*, since it was introduced in awareness that a previous programme failed because of low completion rates. Since the previous failure appears to have arisen in part because of reduction in staff numbers, a more rational response would have been to ensure that a sufficiency of adequately-trained staff to implement the program. The reasons for wanting near-complete screening of the age group in the *B4 School Check*, are different from and far less persuasive than those which might apply in programs for immunization against infectious diseases which might cause an epidemic. (ii) The *B4 School Check* has a single consent form for all tests. I understand that in at least some regions, parents may, if they wish, give consent to only some components, usually vision and hearing, but not the rest. This may vary from one region to another. (iii) The third criterion given above (legislation criminalizing those who break confidence on test results) does not apply to many screening programs with such force, but in the context of HIV testing, was crucial given the high degree of stigmatisation likely to follow public knowledge of a person’s HIV status. The same fear looms large within the mental health service user communities in New Zealand. The perception that public release of data seems inconceivable at present is no safeguard. There is a real possibility of data theft. Moreover government policies and how they reflect public perceptions can change. Once an assessment of a person’s mental health status is written down in official documents, it is hard to rule out the possibility of its being used unfairly against the person concerned at a future date. What often happens in bureaucratic systems is that data systems which initially are subtle, nuanced and non-categorical come to be taken, on the basis of their black-and-white and seemingly-objective nature (however subjective and error-prone they really are) to be more solid as statements about individuals than is justified. Data can then gradually be “morphed” into more simplified categorical systems. It is a real danger.

Population-wide screening makes best sense when: (a) It provides precise and reliable information; (b) That information can be the basis of well-proven action; (c) The intentions of the program are fully explained to the public beforehand. Discussion of the first two points follows later. On the third point, transparency of intended purposes of the SDQ, there appear to be deficiencies.

**[III] Referral Criteria and Referral Streams.** Cut-off scores on the SDQ total score are used to define when children are “concerning”, “some concern”, or

---

39 New Zealand, National Radio, 21.02.2013),

40 *Dominion Post*, 7/8. September, 2013: “Kid’s ear test flaws force wider rechecking”

41 see comments in Parliament of the Minister, Mr Tony Ryall, (Hansard, 13.09. 2011, p 21184)

“normal”. On p.34 of the *Practitioner’s Handbook (Guidelines for interpreting scores)* we read: “Total difficulties scores between 17 and 40: A ‘concerning’ (prev. abnormal) total difficulties score on either of SDQ-P or SDQ-T identifies possible mental health disorders, socio-economic issues or other developmental disorders affecting the child and their family/whanau.” . . . . “Refer a child with ‘concerning’ scores to a paediatrician, a child mental health specialist or the Child and Adolescent Mental Health Services, or Group Special Education, depending on the type of further evaluation and management the child needs. A multidisciplinary team is the ideal referral pathway. The sub-scores may assist in defining the best referral pathway. This is best defined locally, based on the experience and availability of expertise.” It is not clear that these recommendations fulfil criterion (f) of the Guidelines for Health Screening programs (vis: “The health care system is capable of supporting all necessary elements of the screening pathways, including diagnosis, follow-up and programme evaluation”).

Later we read: “If problems are identified, refer children for diagnostic evaluation to determine their eligibility for services. Diagnostic evaluation involves an in-depth assessment with help from the family/whanau and has three outcomes. The diagnosis may be:

- \* of a mental health disorder, so formal intervention is needed
- \* behavioural issues that suggest some form of informal intervention may be helpful such as reading material or a play group.
- \* normal behaviour”<sup>42</sup>

For the less serious scores categorized as “‘some concern’ (borderline)” on the total difficulties score the recommendations are as follows: “If a child has a ‘some concern’ total difficulties score, discuss the child’s strengths and difficulties with their parent and teacher, and consider recommending targeted parenting programmes to support the child and their family/whanau.

One other very significant detail, in a section entitled “Review of ‘Concerning’ Total difficulties score” (p. 34) reads: The cut-off for referral will be reviewed once the *B4 School Check* is rolled out across the country and more information about referrals becomes available. New Zealand norms will be developed once the *B4 School Check* programme has been established and there is sufficient data for a robust study.” In other words, the authors of the *Handbook* admit that at the time of writing, and despite the intended catchment being nation-wide and near-universal, it is experimental. In such circumstances, it would be normal practice to trial the program on a small scale before rolling it out universally. Two other questions follow from this: (i) Has use of the SDQ in *B4 School* been declared as a research study? (ii) Has it been submitted for approval to an authorised ethics committee?

## **E: Critique of the SDQ.**

**[I] Introduction.** Before going into detail, the reasons given for using the SDQ rather than some other instrument should be made clear: that it is “widely used and

---

42 Note that while the *Practitioner’s Handbook* recommends avoiding the word ‘abnormal’ the word ‘normal’ still creeps in surreptitiously.

accepted by people working in child health, development and education; quick to use and easy to score; valid, sensitive and specific; free to download from the *Youth in Mind* website.<sup>43</sup> Readers are reminded of the National Health Committee Guidelines, on health screening programs that there should be “a suitable test for the condition”, and in the *Practitioners Handbook for B4 School* (p.27), that “tools to identify child development/behaviour should be simple to administer and score<sup>44</sup>. They should have sensitivity and specificity above 70 percent.” This figure is not mentioned in the Guidelines on Health Screening programs; and it is right that no specific figure should be stated: When the condition to be detected in the screening program is relatively rare, to reach an admirably low percentage of false positives in the remaining vast majority may still mean that false positives greatly outnumber true positives in the minority group. This might be a strong argument against use of the screening system.

**[II] *The Overall Style and Mode of Delivery of the “Strengths and Difficulties Questionnaire”.*** Since the SDQ is to be used for individual assessment, and words to be used by practitioners is specified with some precision, it appears to assume that words have similar significance for all respondents, regardless of their cultural, national, educational or social level, or preferred language. This assumption is clearly invalid, and a good clinician always adjusts his/her language to meet the capacity of each patient. In addition, based on their differing experiences, parents (though perhaps not teachers) may differ in their views of what is normal or abnormal in a child. This suggests that different parent’s responses on the SDQ may not mean the same thing. Thus, while a scale like this may be valuable for research comparing aggregate scores between groups<sup>45</sup> or across time, on both counts it is less suitable for individual assessment, and as a screening tool.

It might then be predicted that there be poor correlation, and perhaps systematic differences, between ratings based on reports of teachers, parents and self-report, or between any of these and clinicians’ diagnoses. Comparisons, including one from mainland China, give correlation coefficients for the four main scales and total scores between different raters. Between parent and teacher reports coefficients range from 0.23 to 0.54, between teacher and youth reports from 0.19 to 0.45, and between parent and self-reports from 0.14 to 0.56<sup>46</sup>. Correlation between caregiver

---

43 <http://www.sdqinfo.com>

44 Tsiantis,J, Smith,M, Dragoness,T, Cox,A. (2000) Early mental health promotion in children through primary health care services: A Multi-Centre Implementation. *International Journal of Mental Health Promotion*, 2 (3), 5-17.

45 As in the following study: Goodman A. Goodman R. (2011) Population mean scores predict child mental disorder rates: validating SDQ prevalence estimators in Britain. *Journal of Child Psychology & Psychiatry & Allied Disciplines*. 52(1):100-108.

46 Koskelainen,M, Sournder,A, Kaljonen,A. (2000) The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. *European Child & Adolescent Psychiatry*, 9, 277-284; Goodman,R (2001) Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40, 1337-1345; Muris,P, Meesters,C, van den Berg,F. (2003) The Strengths and Difficulties Questionnaire (SDQ): Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child & Adolescent Psychiatry*, 12, 1-8; Widenfelt, BM, Goedhart,AW, Treffers, PDA, Goodman, R. (2003) Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child & Adolescent Psychiatry*, 12, 281-289; Du,Y, Kou,J, Coghill,D. (2008) The validity, reliability

and youth reports tended to be higher for externalizing disorders (attention deficit disorder, rule-breaking behaviour, aggression) than for internalising ones (anxiety, depression and somatic complaints)<sup>47</sup>. More useful than correlation coefficients is the proportion of variance in one measure accounted for by variation in the other, computed as the square of the correlation coefficient. Values just quoted then imply that 2-29% of variance in one type of report depends on something in common with scores on other type of report. Regardless of detail, by far the largest determinants of scores given by one type of reporter differs from that for another, for the same child.

Systematic differences in SDQ scores obtained from different types of reporter are documented in a study from Quebec<sup>48</sup>. Amongst children aged 6-8 years from the general population, internalising problems were identified in 11.3% of child reports, and 18.3% of parent reports; externalising problems were identified in 6.5% of child reports, 8.25% of parent reports, and 10.8% of teacher reports. At age 9-11 years, internalizing problems were identified in 8.1% of child reports, and 15.4% of parent reports. Externalising problems were identified in 8.5% of child reports, 6.1% of parent reports, and 8.9% of teacher reports. At age 12-14 years internalising problems were identified in 14.2% of reports from adolescents, and in 13.7% of parent reports; externalising ones in 4.4% of reports from adolescents, and 5.8% of parents' reports. In the Chinese study referred to above, most problems were reported at more severe levels by parents than by teachers, although the difference was less in older age groups. Overall, the biggest differences between reporters are seen for the youngest age group, and for internalising disorders.

Other concerns are that wording<sup>49</sup> or cut-off scores<sup>50</sup> of the SDQ determined in majority populations cannot be equated with those for indigenous or culturally and linguistically diverse communities, and that cut-offs cannot be equated from country to country<sup>51</sup>, or from one social group to another in a country. Moreover, correlates

---

and normative scores of the parent, teacher and self report versions of the Strengths and Difficulties Questionnaire in China. *Child and Adolescent Psychiatry and Mental Health* 2,8. Accessed at: <http://www.capmh.com/content/2/18>;

47 Penney,SR, Skilling, TA (2012) Moderators of informant agreement in the assessment of adolescent psychopathology: Extension to a forensic sample. *Psychological Assessment*, 24, 386-401.

48 Bergeron,L, Valla,J-P, Breton,J-J, Gaudet,N, Berthiaume,C, Lambert,J, St-Georges,M, Smolla,N. (2000) Correlates of mental disorders in the Quebec General Population of 6-14-year olds. *Journal of Abnormal Child Psychology*, 28, 47-62.

49 Williamson,A, Redman,S, Dadds,M, Daniels,J, D'Este,C, Raphael,C, Eades,S, Skinner,T (2010) Acceptability of an emotional and behavioural screening tool, for children in aboriginal community controlled health services in Urban NSW. *Australian and New Zealand Journal of Psychiatry*, 44, 894-900.

50 Priest N. Baxter J. Hayes L. (2012) Social and emotional outcomes of Australian children from Indigenous and culturally and linguistically diverse backgrounds. *Australian & New Zealand Journal of Public Health*, 36(2):183-190; Shojaei,T, Wazana,A, Pitrou,I, Kovess,V. (2009) The strengths and difficulties questionnaire: validation study in French school-aged children and cross-cultural comparisons. *Social Psychiatry & Psychiatric Epidemiology*: 44, 740-747.

51 Goodman A. Heiervang E. Fleitlich-Bilyk B. Alyahri A. Patel V. Mullick MS. Slobodskaya H. Dos Santos DN. Goodman R. (2012) Cross-national differences in questionnaires do not necessarily reflect comparable differences in disorder prevalence. *Social Psychiatry & Psychiatric Epidemiology*, 47(8):1321-1331.



of behavioural and emotional problems differ if assessed in different situations<sup>52</sup>. This suggests that situational factors play a part in determining symptoms<sup>53</sup>. To use concepts of “diagnosis” and “disorder” may thus be inappropriate.

**[III] Validation of the SDQ in the Target Age-Group.** Many papers have explored the reliability and validity of the SDQ, but few if any (I have found none!) refer just to pre-schoolers (age 4-5)<sup>54</sup>. The only paper I found which singles out the youngest age group<sup>55</sup> uses the German version of the SDQ to assess 3-17 year-olds, and reports that conduct problems were significantly worse in 3-6 year-olds, than in older age groups. A Japanese study<sup>56</sup> found that ratings on conduct disorder and hyperactivity/intention subscales peaked at age 7-9 years, but fell in older ages to levels below that of 4-6 year-olds. A Chinese study<sup>57</sup> comparing years 3-10 with 11-17, found a decrease with age in SDQ-rated emotional and hyperactivity symptoms, with little change in symptoms of conduct disorder. An earlier Finnish study<sup>58</sup> had also found that problems identified by the SDQ in youngsters under 12 years tended to normalise by teenage years. Thus, while there may be differences across cultures, and for different types of problem, such problems (including conduct disorder) are in part ones of growing up, as much as actual disorders.

**[IV] Biased Coverage of Child Mental Health Problems by the SDQ.** A possible criticism of the SDQ to screen for incipient mental disorders in young people is that it is biased in favour of some issues, but fails to cover the full range of child psychiatric problems. Perhaps because of its brevity the SDQ may be unable to detect several significant problems, some of which are rare, others more common. To discover if this is the case, the subsections below give details on: (i) detection rates and prevalence of problems identified with SDQ and its subscales; (ii) relative proportions of diagnoses of persons picked up using the SDQ; (iii) the prevalence of child mental health problems in general population health surveys by other methods.

---

52 Szatmari,P, Offord,DR, Siegel,LS, Finlayson,MAJ and Tuff,L. (1990) The clinical significance of neurocognitive impairments among children with psychiatric disorders: diagnosis and situational specificity. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 31, 287-299.

53 Aschenbach,TM, McConaughy,SH, Howell,CT (1987) Child/adolescent behavioural and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101, 213-232.

54 I understand that such an evaluation is currently being carried out within Auckland University of Technology. The results are expected late in 2013, four years after widespread roll-out of *B4 School* with the SDQ as one of its components.

55 Hölling,H, Kurth,B-M, Rothenberger,A, Becker,A, Schlack,R. (2008) Assessing psychopathological problems of children and adolescents from 3 to 17 years in a nationwide representative sample: results of the German health interview and examination survey for children and adolescents. *European Child & Adolescent Psychiatry*, 17, suppl, 1, 34-41.

56 Matsuishi,T, Nagano,M, Araki, Y, Tanaka,Y, Iwasaki,M, Yamashita,Y, Nagamitsu,S, Iizuka,C, Ohya,T, Shibuya,A, Hara,M, Matsuada,K, Tsuda,A, Kakuma,T. (2008) Scale properties of the Japanese version of the Strengths and Difficulties Questionnaire (SDQ): A study of infant and school children in community samples. *Brain & Development*, 30, 410-415.

57 Du,Y, Kou,J, Coghill,D. (2008) The validity, reliability and normative scores of the parent, teacher and self report versions of the Strengths and Difficulties Questionnaire in China. *Child and Adolescent Psychiatry and Mental Health* 2,8. Accessed at: <http://www.capmh.com/content/2/18>

58 Koskelainen,M, Sournder,A, Kaljonen,A. (2000) The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. *European Child & Adolescent Psychiatry*, 9, 277-284.

(i) *Coverage and Prevalence as determined by SDQ subscales.* The types of problem identified by the SDQ can be gauged from statistics on those who reach thresholds (for “possible” or “probable”) on the four subscales of the SDQ. Results for community samples, as percentages (sometimes split by gender [boy/girl]), are presented in Table I. Cut-offs are those for “SDQ-Probable”, except in the German study of Hölling et al (2008), where the “SDQ-Possible” cut-offs are used, which would elevate the percentages overall. In these data, there is a regular tendency for conduct and hyperactivity disorders, as detected by the SDQ to be more common than emotional problems detected with this instrument.

**TABLE I: Prevalence estimates\* of Disorders Identified by the SDQ**

<i>Study</i> <sup>59</sup> , <i>Country, age</i>	<i>n</i>	<i>Conduct Disorder</i>	<i>Hyperactivity Inattention</i>	<i>Emotional Problems</i>	<i>Peer Problems</i>	<i>Total/Any Problems</i>
Goodman et al 2000, UK	7894(PT)	3.9	1.9	1.3	NA	9.2
Ronning et al 2004, Norway 11-16y.	4167(S)	9.8/4.1	11.1/8.8	6.5/12.5	10.4/7.2	10.4/7.6
Hölling et al 2008, Germany 3-17 y	14,478(P)	17.6/11.9	18.2/9.3	15.5/17.2	13.1/9.9	22.0/14.8
Elhamid et al, 2009, Egypt 6-12 y	1167(T) 1077(P)	27.7 25.3	16.6 10.7	15.3 21.2	23.7 33.8	34.7 20.6

\* Percentages, boy/girl, when available; T: Teacher report; P: Parent report; S: Self-report

(ii) *The relative proportion of diagnoses detected by SDQ screens, and the sensitivity and selectivity of such screening.* Two community surveys provide statistics on official diagnoses made by clinicians (or by parents using criteria more specific than the SDQ) for young people initially identified as having problems with the SDQ. Conduct and hyperactivity are more common than the anxiety/depression constellation, as shown in Table II. A smaller study from Australia<sup>60</sup> was based on 130 new referrals (aged 4-15 years) to a child and adolescent mental health clinic in

---

59 References as follows: Goodman,R, Ford,T, Simmons,H, Gatward,R, Meltzer,H. (2000) Using the strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, 177, 534-539; Ronning,JA, Handegaard, BH, Sourander,A, Morch,W-T. (2004) The Strengths and Difficulties Self-Report Questionnaire as a screening instrument in Norwegian community samples. *European Child & Adolescent Psychiatry*, 13, 73-82; Hölling,H, Kurth,B-M, Rotherberger,A, Becker,A, Schlack,R. (2008) Assessing psychopathological problems of children and adolescents from 3 to 17 years in a nationwide representative sample: results of the German health interview and examination survey for children and adolescents (KiGGS). *European Child & Adolescent Psychiatry*, 17, 34-41; Elhamid,AA, Howe,A, Reading,R. (2009) Prevalence of emotional and behavioural problems among 6-12 year old children in Egypt. *Social Psychiatry & Psychiatric Epidemiology*, 44, 8-14;

Melbourne. SDQ reports from parents (n=130), teachers (n=101) and self-reports (n=38), were compared with diagnoses from the clinical team, and an independent clinician. The SDQ identified 93% of conduct disorder cases diagnosed clinically, 56% of hyperactivity disorder cases, but only 36% of emotional disorders. In other words, sensitivity for detecting conduct disorder (and to lesser extent, hyperactivity) was better than that for detecting emotional problems.

**TABLE II: Prevalence estimates of some clinical diagnoses detected using the SDQ**

<i>Study<sup>61</sup>, country age</i>	<i>n</i>	<i>Diagnoses</i>				
		1.	2.	3.	4.	5.
Goodman et al, 2000, UK, 5-10 y.	4776	3.2	1.84	1.38	0.18	0.39
Ravens-Sieberer et al, 2008, Germany, 7-10y.	948	16.1/14.8	6.26/2.6	12.3/16.5	11.4/10.0	-

Key to diagnoses: 1: Conduct disorder; 2: ADHD; 3: Anxiety; 4: Depression; 5: Less common diagnoses;

(iii) *Prevalence of child psychiatric disorders documented in other ways.* Few data on prevalence of a variety of different types of childhood psychiatric disorder are available from community surveys prior to the development of the SDQ, and, since the year 2000, most studies of this used the SDQ. In a study in 1994 from Spain<sup>62</sup>, ADHD was the most common DSM-III-R diagnosis in eight-year olds, the prevalence falling at older ages. Disturbed behaviour was relatively uncommon (and probably was so at earlier ages), but increased with age. A few community surveys used a broader categorization, documenting prevalence of “externalising” versus “internalizing” problems. The former are likely to include the most cases of hyperactivity and conduct disorder, the latter most of the emotional problems scored in the SDQ. In the survey from Quebec cited above, of internalizing problems were considerably more common than externalising ones. This corroborates an earlier

60 Mathai,J, Anderson,P, Bourne,A. (2004) Comparing psychiatric diagnoses generated by the Strengths and Difficulties Questionnaire with diagnoses made by clinicians. *Australian and New Zealand Journal of Psychiatrists*, 38, 639-643.

61 Goodman,R, Ford,T, Simmons,H, Gatward,R, Meltzer,H. (2000) Using the strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, 177, 534-539; Ravens-Sieberer,U, Wille,N, Erhart,M, Bettge,S, Wittchen,H-U, Rothernberger,A, Herpertz-Dahlmann,B, Resch,F, Bullinger,M, Schukte-Markwort,M, Döpfer,M (2008) Prevalence of mental health problems among children and adolescents in Germany: results of the BELLA study within the National Health Interview and Examination Survey. *European Child & Adolescent Psychiatry*, 17, 22-33.

62 Gómez-Beneyto,M, Bonet,A, Catala,MA, Puche,E, Vila,V. (1994) Prevalence of mental disorders among children in Valencia, Spain. (1994) *Acta Psychiatrica Scandinavica*, 89, 352-357.

report<sup>63</sup> based on a survey of 664 outpatients seen in a child psychiatry clinic. While not a general population survey, it showed that emotional disorders (present in 42.2% of boys and 58.6% of girls) were more common than conduct disorder (13.5% and 4.0%) or mixed conduct/emotional disorders (18.5% and 11.7%). Disturbed relationships were quite common (26.0/25.7%). The most common internalising problems in childhood and adolescence were anxiety disorders, with 2-17%, and 2.5-5% (respectively) meeting diagnostic criteria in any age group<sup>64</sup>. A 2007 study from the USA<sup>65</sup> of a random population sample (actually from a twin register) of 814 8-year-olds found the most common DSM-IV diagnosis in boys was ADHD (18.3%), with externalising disorders (oppositional/defiant disorder plus conduct disorder making up 12.5%, and internalising disorders (generalised anxiety disorder, separation anxiety disorder, and major depression) making up 5%. In girls the percentages were 6.8% (ADHD), 8.0% (externalising disorders), and 7.8% (internalising disorders).

(iv) *Conclusions*: From several sources, it appears that the SDQ detects most cases of conduct disorder/hyperactivity, but fails to detect other child psychiatric problems, especially internalising or emotional problems. Thus, even given that it is for screening rather than diagnosis, the SDQ may not be the best broad-screening tool for early child mental health. The SDQ not only detects more abnormalities in the conduct disorder and hyperactivity domains than in other areas, but in the former appears to be more reliable (i.e. has a higher correlation with supposedly authentic diagnoses). Thus in a Swedish study<sup>66</sup>, specificity figures for identifying conduct and hyperactivity disorders were 81.7% and 79.1% respectively, compared to 76.8% for emotional difficulties and 72.1% for peer problems). Likewise in the Australian study correlation coefficients between SDQ assessments and clinical diagnoses were 0.56 for conduct disorder, 0.44 for hyperactivity, but only 0.39 for emotional problems. For diagnoses from an independent clinician the values were 0.55 for conduct disorder, 0.65 for hyperactivity, and 0.45 for emotional problems.

The focus on conduct disorder in a pre-school check said to be on *health issues* is somewhat anomalous. In Ministry of Health data on youth mental health (ages 2-14

---

63 Kosky,R, McAlpine,I, Silburn,S, Richmond,J. (1985) A survey of child psychiatry outpatients. 1. Clinical and demographic characteristics. *Australian and New Zealand Journal of Psychiatry*, 19, 158-166.

64 Grills-Tauechel,A., Ollendick,TH (2007) Introduction to special issue: Developments in the etiology and psychosocial treatments of anxiety disorders in children and adolescents. *Clinical Child and Family Psychology Review*, 10, 197-198; Rapee,RM, Schiering,CA, Hudson,JL (2009) Anxiety disorders during childhood and adolescence: Origins and treatment. *Annual Review of Clinical Psychology*, 5, 311-341; Zahn-Waxler,C, Klimes-Dougan,B, Slattery,MJ (2000) Internalising problems of childhood and adolescence: Prospects, pitfalls and progress in understanding the development of anxiety and depression. *Development & Psychopathology*, 12, 443-466.

65 Lemery-Chalfant,K, Schreiber,JE, Schmidt,NL, Van Hulle,CA, Essex,MJ, Goldsmith,HH. (2007) Assessing internalising, externalizing, and attention problems in young children: validation of the MacArthur HBQ. *Journal of the American Academy of Child & Adolescent Psychiatry*, 46, 1315-1323.

66 Malmberg,M, Rydell,A-M, Smedje,H. (2003) Validity of the Swedish version of the Strengths and Difficulties Questionnaire (SDQ-Swe). *Nordish Journal of Psychiatry*, 57, 357-363.

years)<sup>67</sup>, categories for which statistics are given include depression, anxiety and ADD/ADHD. Conduct disorder is not mentioned, although both DSM-IV and ICD-10 define this diagnosis. *B4 School* originated in the Ministry of Health, and most of its screening tests fit well into this Ministry's agenda. However, this anomaly adds weight to the conjecture that the drivers of the behavioural testing component are elsewhere, especially in the Ministry of Social Development, and its Interagency Plan for Conduct Disorder/Severe Antisocial Behaviour (input to which came from Ministries of Education, Justice, Social Development, the High and Complex Needs Unit, as well as the Ministry of Health). The ambivalence on whether behavioural testing is viewed within medical, educational or other paradigms, while using an instrument definitely originating in medical/psychiatric circles, betrays this complex provenance. Perhaps this should have been made clear in presenting *B4 School* to the public. In addition, for a health screening program, one would have thought that the instrument used would be effective in at least detecting, if not identifying the most common mental health problems; but the SDQ is not effective. One therefore suspects that the real emphasis in this part of *B4 School* is to protect the public (and the public purse), rather than to safeguard mental health of individual children. If so, contrary to Margaret Thatcher's line the guideline seems to have switched to "There is no such thing as the individual, only society!"

**[V] Ability of the SDQ to Discriminate Different Types of Problem.**

The *Practitioner's Handbook* for *B4 School* clearly indicates that scores for the separate subscales of SDQ are relatively valid measures to identify separately the problems addressed by each subscale. This is also clear in the *YouthinMind* website, linked to *B4 School*, and operates in many languages across many countries as an international youth mental health industry. However, one should ask whether the SDQ really has the ability to differentiate problems corresponding to each subscale not only from normal, but more importantly, from each other?

Research studies question this. One way to examine discriminating ability of the different subscales is to compute how well subscale scores correlate with each other, low correlation coefficients implying relative independence of subscales. In several studies<sup>68</sup> correlations between scores for Hyperactivity and Conduct Problems were high ( $r=0.39-0.84$ ), as were those between Peer Problems and Emotional Problems ( $r=0.62-0.72$ ). Those between hyperactivity/inattention or conduct disorder and emotional problems were much lower ( $0.19-0.38$ ).

---

67 Ministry of Health (2011) Health of New Zealand Child, 2011/12, Key findings of the New Zealand Health Survey, Section 2: Health Status and Conditions. Ministry of Health, Wellington.

68 Goodman,R (2001) Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40, 1337-45; Van Roy,B, Veenstra,M, Clench-Aas,J. (2008) Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry*, 49, 1304-1312; Du,Y, Kou,J, Coghill,D. (2008) The validity, reliability and normative scores of the parent, teacher and self report versions of the Strengths and Difficulties Questionnaire in China. *Child and Adolescent Psychiatry and Mental Health* 2,8. Accessed at: <http://www.capmh.com/content/2/18>; Sanne,B, Torsheim,T, Heiervang,E, Stormack,KM (2009) The Strengths and Difficulties Questionnaire in the Bergen Child Study: A Conceptually and methodically Motivated Structural Analysis. *Psychological Assessment* 21, 352-364.

A more powerful method is to document discrepancies between determinations on SDQ subscales, and those made in other ways (including, but not limited to clinical diagnoses). In an early study with the SDQ<sup>69</sup>, one of several diagnoses was available for the cohort. The SDQ assignment to a disorder was generally more common than the assignment to a corresponding diagnosis. Major discrepancies were found for diagnoses of anxiety or depression (where the commonest SDQ designation was “conduct disorder” [47.8% of cases], not “emotional disorder” [34.4% of cases]). A Dutch study<sup>70</sup> reports correlations between SDQ subscale scores as rated by parents and teachers, and scores on the Child Behaviour Checklist (CBCL) as rated by parents. For the SDQ “probable” determination of an emotional problem, the highest CBCL correlation was with anxiety/depression ( $r=0.70$  and  $0.65$  for parent and teacher ratings respectively), and for “internalizing problems” ( $r=0.70/0.64$ ). For “withdrawn” the  $r$ -values were  $0.52/0.48$ , for “social problems”:  $0.51/0.39$  and for “attention problems”  $0.45/0.50$ . For SDQ “probable” scores for Conduct Disorder, the highest correlation was with “externalising” ( $r=0.72/0.62$ ), “aggressive” ( $r=0.71/0.57$ ), “delinquent” ( $r=0.56/0.58$ ). For the SDQ “probable” score for Hyperactivity/Inattention, the highest correlation was with “attention problems” ( $r=0.78/0.66$ ), “aggressive” ( $r=0.51/0.48$ ), “externalizing” ( $r=0.50/0.50$ ) and “Peer: Social Problems” ( $0.51/0.41$ ). For SDQ-total difficulties “probable” determination, the strongest correlations were with attention problems ( $0.76/0.73$ ) “aggressive” ( $0.64/0.51$ ), “externalizing” ( $0.63/0.53$ ), “social problems” ( $0.62/0.49$ ), internalizing ( $0.56/0.49$ ) and “anxiety/depression” ( $0.55/0.45$ ). In this study, the highest correlations *were* with the most appropriate clinical diagnosis. However, the correlations were far from perfect, remembering that a correlation coefficient of  $0.7$  implies that  $\sim 50\%$  of variance in one measure is not accounted for by variance in the other. Generally the SDQ was better for recognizing externalizing than internalizing problems. A more recent study<sup>71</sup> focused on use of the SDQ for recognizing ADHD also found it to be both highly sensitive and highly selective for the hyperactive variety, but far less so for the inattentive variety.

A large study from Texas<sup>72</sup> was unusual in that the age range of the children on which it focused (low achievers, aged  $6.1\pm 0.65$  years) was close to that intended for *B4 School*. After a thorough examination of the psychometric properties of the SDQ, using data from parents, teachers and peers (nominated classmates), they conclude “This instrument would be of particular value to school psychologists involved in assessing the effectiveness of universal prevention programs in improving mental health and decreasing symptoms of emotional and behavioural

---

69 Goodman,R, Ford,T, Simmons,H, Gatward,R, Meltzer,H. (2000) Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, 177, 534-539.

70 Widenfelt, BM, Goedhart,AW, Treffers, PDA, Goodman, R. (2003) Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child & Adolescent Psychiatry*, 12, 281-289.

71 Illebo,AK, Posserud,M-B, Heiervang,E, Gillbeeg,C, Obel,C. (2011) Screening for the attention deficit hyperactivity disorder phenotype using the strength and difficulties questionnaire. *European Child & Adolescent Psychiatry*, 20, 451-458.

72 Hill,CR and Hughes,JN (2007) An examination of the convergent and discriminant validity of the Strengths and Difficulties Questionnaire. *School Psychology Quarterly*, 22, 380-406.

problems. However, the SDQs's brevity comes at the expense of its being able to discriminate among different types of problems."

**[VI] Specificity, and the Problem of False Positives:**

(i) *Evidence.* Attention has been drawn to the low level of agreement for child and adolescent disorders between practitioners' routine diagnoses, and standardised research diagnoses<sup>73</sup>. In recent years, Allen Frances also warns about the dangers of diagnostic inflation<sup>74</sup>, and the increasing difficulty of diagnosis with decreasing age. However, when SDQ assessments are compared with diagnoses, it is important to consider the specificity, or, as the reverse of this, the proportion of false positives – those rated by the SDQ as “probable”, but who, on specialist assessment, do not meet criteria for any diagnosis. In one study from Goodman's group<sup>75</sup>, the SDQ category “disorder unlikely” proved to be correct in 98.8% of cases, the “disorder possible” category led to a confirmed diagnosis in 10.8% of cases, while the “disorder probable” category was confirmed in 52.7% of cases. In other words, nearly half of those where any disorder is judged probable by SDQ (in a community sample) did not merit any diagnosis. In another study from Goodman's group<sup>76</sup>, with cohorts from London and Dhaka, it was found that there were more false positives than negatives, so the SDQ could be said to be “over-inclusive”. This study was conducted under research conditions in specialty clinics. The authors suggest that, even there, the SDQ is useful mainly for drawing attention to possible problems, rather than defining what they are. They suggest that it might not be good enough with community samples, such as those involved in *B4 School*.

Later studies reach similar conclusions. A Swedish study<sup>77</sup> found that the specificity of the SDQ ranged from 55.8% to 81.7% on its various subscales (and 82.7% on “total difficulties”), implying false positives in 17.3% to 44.2% of cases. In an Australian study<sup>78</sup>, based on a community sample of 1359 children, aged 4-9 years, the prevalence of DSM-IV diagnoses was reported for children whose SDQ total or subscale scores fell in the highest 10% (a criterion similar to the “SDQ-probable” threshold scores). Only 36% of those whose total SDQ scores fell in this percentile met criteria for “any diagnosis”, only 9.8% of those assessed “probable” on the conduct disorder subscale met those for DSM-IV Conduct Disorder, 20.8% of those “probable” on the hyperactivity subscale met criteria for “any ADHD”, 17.5% of those “probable” on the Emotional Problems subscale met criteria for

---

73 Jensen,AL, Weisz, JR (2002) Assessing match and mismatch between practitioner-generated and standardised interview-generated diagnoses for clinic-referred children and adolescents. *Journal of Consulting and Clinical Psychology*, 70, 158-168.

74 Batstra,L, Frances,A. (2012) Diagnostic inflation. Causes and a suggested cure. *Journal of Nervous and Mental Disease*, 200, 474-479.

75 Goodman,R, Ford,T, Simmons,H, Gatward,R, Meltzer,H. (2000) Using the strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, 177, 534-539.

76 Goodman,R, Renfrew,D, Mullick,M, (2000) Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child & Adolescent Psychiatry*, 9, 129-134.

77 Malmberg,M, Rydell,A-M, Smedje,H. (2003) Validity of the Swedish version of the Strengths and Difficulties Questionnaire (SDQ-Swe). *Nordish Journal of Psychiatry*, 57, 357-363.

78 Hawes,DJ, and Dadds,MR (2004) Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Australian and New Zealand Journal of Psychiatry*, 38,644-651.

“Internalising Disorder”, and 26.5% of those on the Peer Problems, for “any diagnosis”. Clearly, over half of those identified as having a problem by the SDQ fail to meet diagnostic criteria for any disorder, and for specific disorders, subscale scores were even worse in identifying corresponding clinical conditions.

Another recent study<sup>79</sup> assessed 328 children from low-income but otherwise representative families, at four stages from preschool up to Grade 5 (ages 9-11 y). Various scales were used to rate externalising or internalising symptoms, singly or in combination. The children were then assigned to one of four groups, according to the longitudinal course of their symptoms. (i) Never any symptoms (58.8%); (ii) Isolated symptoms (21.6%): These could be either internalising or externalising symptoms, with subsequent return to low symptom level; (iii) Recurrent symptoms but without combination of internalising and externalising symptoms (10.1%); (iv) Recurrent symptoms with externalising and internalising symptoms in combination (9.5%). In terms of Grade 5 outcomes, the most relevant data were on use of mental health services, which could be either those based in schools, or in specialty clinics. In the four groups, the percentages needing such services were (respectively): 10.4%/3.1%; 21.1%/11.4%; 21.2%/21.2%; 48.4%/36.7%. In other words, by collecting data from kindergarten and several subsequent stages, one can predict outcomes needing mental health services at grade 5 with 35-50% specificity for the fourth (most severe) group, and, for the next most-severe category, with ~20% specificity. The authors comment: “The findings indicate that children most likely to develop recurrent co-morbid symptoms can be identified quite accurately by the end of grade 1 using a relatively low-cost universal screening procedure based on mother and teacher questionnaire reports of children’s mental health symptoms during the transition to primary school. Only the small subgroup of children who screen positive would then be referred for more costly expert diagnosis and intervention.” It is hard to find support for such a strong conclusion in the evidence these authors provide. Moreover, this was a research study, not a universal roll-out, and the level of prediction achieved required more effort than a single preschool SDQ test. In this context, in the *Practitioner’s Handbook for B4 School*, (p.2) we read: “Assessment is not a static process. Effective assessment cannot be a one-off action, but is an on-going process. The assessment process is the beginning of the therapeutic relationship with the child and family/whanau.” However, as already suggested, current implementation of the SDQ does not involve repeat assessment over several years, nor was it intended to do so. It is not clear that this would ever be feasible for mass screening of an entire age group. These commendable guidelines probably exceed resources likely to be available to implement them.

(ii) *The balance between specificity and sensitivity*: Any health or behaviour screening program has to pay careful attention to both specificity and sensitivity, or, conversely, the proportion of errors to be classed respectively as false positives and false negatives. Generally these two constraints are mutually opposed: The greater attention paid to achieving one, the less can be achieved with respect to the other.

---

79 Essex,MJ, Kraemer,HC, Slattery,MJ, Burk,LR, Boyce,WT, Woodward,HR, Kupfer,DJ (2009) Screening for childhood mental health problems: outcomes and early identification. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 50, 562-570.



One or other may need to be given priority: If what is being screened for is urgent and serious, avoiding false negatives may take precedence; if there are serious detrimental effects of false positives, avoiding *them* may have a higher priority. If a screening process is easy, cheap and innocuous, and what is being screened for is potentially a serious and expensive health risk, a false-positive rate of 50% or more may be acceptable. There is no correct answer to this; it is often a matter of controversy. For instance, the usual test for risk of prostate cancer in men (the PSA blood test) is controversial: In the USA it is recommended by the FDA for men over 50 years, but not by the Preventive Services Task Force, nor in Britain by the NHS, where, on the basis of statistics, it is argued that “12 to 48 men would be treated unnecessarily for a disease that would never cause them problems”<sup>80</sup>.

When a scale is used in screening which, like the SDQ, gives a quantitative or semi-quantitative measure, the cut-off point at which intervention is recommended is adjustable. If the detection threshold is lowered, the rate of false negatives will fall and that of false positives will rise. The same opposed requirements may also be in play in deciding which scale to use for a particular detection tasks. In the case of early detection of incipient conduct disorder, one study<sup>81</sup> compared the Child Behavior Checklist (CBCL) and the Behavior Problem Checklist (BPC). The CBCL, as usually used, had higher specificity (few false positives), and lower sensitivity (missing many real cases) than the BPC.

(iii) *Comment on use of the SDQ*: The rate of false positive identifications made by the SDQ is very high. Several further points should be borne in mind here.

In child psychiatry the question often is “how much better is an ‘authentic’ diagnosis compared with one delivered by the SDQ or any other screening tool.” A recent study<sup>82</sup> makes this question very pertinent: In a cohort of children from a paediatric out-patient clinic, 60/307 were identified with the SDQ as having “probable” psychiatric disorders. Only 25% of these cases had been identified by doctors at the clinic, although 33 of the 60 had already received professional help of some sort (often from more than one source). One should then ask: What is the “gold standard” for childhood psychiatric diagnoses? It is easily argued that there is no highly credible standard of diagnostic validity, to which other systems can be compared. This is one sign of the insecure status of diagnosis in psychiatry as a whole, especially in young persons. Another indication comes from a study<sup>83</sup> showing high rates of co-morbidity for the diagnosis of “conduct disorder”, that is other disorders occurring at higher-than-expected rates in those who have this as a primary diagnosis. In this case co-morbid diagnoses include (not surprisingly) oppositional/defiant disorder, but also depressive disorders, attention

---

80 Ian Sample (2013) Scientists unravel genetic causes of prostate, breast and ovarian cancer. *Guardian*, 27.03.2013.

81 Kazdin EA, Heidish,IF, (1984) Convergence of clinically derived diagnoses and parent checklists among inpatient children. *Journal of Abnormal Child Psychology*, 12, 421-435.

82 Glazebrook,C, Hollis,H, Heussler, Goodman,R, Coates,L. (2003) Detecting emotional and behavioural problems in paediatric clinics. *Child: Care, Health and Development*, 29, 141-149.

83 Lambert, EW, Wahler,R, Andrade,RG, Bickman,L (2001) Looking for the disorder in Conduct Disorder. *Journal of Abnormal Psychology*, 110, 110-123.

problems, dysthymia, and anxiety problems. It has also come to be recognised by some researchers<sup>84</sup> that conduct disorder is not a single entity, but a mix of different problems which, superficially, look similar. Generally, diagnostic criteria based primarily on overt behaviour without looking more deeply at underlying cognitive, perceptual or emotional roots of behaviour are likely to be simplistic (perhaps driven by long-outdated behaviourist concepts). Throughout psychiatry, the profession is hampered by lack of precise concepts. Without precise concepts, precise reasoning, precise statistical analysis, and of course precise intervention become impossible.

Perhaps more important, in the mental health area there is always hidden concern, strongly felt by many service users, that diagnostic labels may be used to stigmatise and discriminate. The intentions of those who plan and implement a screening program may be entirely honourable; but once data is collected and codified in seemingly objective manner, its future uses may change, as policies of health authorities, and their political masters change. All qualifications, the “ifs and buts” which the original collectors of data were aware of, may by then be forgotten by those who implement different policies, on the basis of past data. In modern health research, a routine requirement on which ethics committees insist, is that the future of personal data be specified. Often a date should be specified by which time the data is shredded. However, there appears to have been no ethical scrutiny of the SDQ. The fact that it was introduced with no clear referral pathway for those identified as having problems, leaves one suspecting that the real motive may have been to collect data for use at a future date, for an unspecified purpose. This is a matter of serious concern<sup>85</sup>.

In this context, the high rate of false positives associated with the SDQ both for specific problems and for “any problem” is worrying, for several reasons: It is worrying for present or potential future users of services for reasons just given. It is worrying for delivery of mental health care:- As pointed out by Allen Frances, for lack of resources, services often cannot *now* deliver adequate care for what are known to be serious problems. Why then use a screening system which “over-predicts” to a considerable extent? It is worrying because it is a drastic departure from guidelines mentioned at the start of this section, that specificity of screening should reach 70%. Clearly the SDQ does not achieve this, even for detecting “any problem”, let alone identifying specific problems (one of the SDQ’s stated aims). An additional source of concern is that that this has not led to more hesitation about use of

---

84 Rutter, M (1995) Relationship between mental disorders in childhood and adulthood. *Acta Psychiatrica Scandinavica*, 91, 73-85.

85 In this context, the Official Information Act (1982), does sometimes permit access to personal information. However, Section 24/3A states: “Where any person is given access to personal information under this section, that person shall be advised of that person's right, under [section 26](#), to request the correction of that information.” (see also Section 26, on “Correction of Information”.)

the SDQ, not least in roll-out of *B4 School* in New Zealand, given that the high rate of false positives was recognised by those developing the SDQ by the year 2000.

One may also ask about the air of accuracy, and ability to discriminate between different sorts of problems conveyed (or claimed) in presenting the SDQ to participants. Claims of ability to discriminate have been discussed above. In conveying the idea that the SDQ has a degree of accuracy, the total-score for a “concerning” (or “disorder probable”) assessment is given as a range from 17 to 40 for parent reports and from 16 to 40 for teacher reports. For the “some concern” (or “disorder possible”) category, the corresponding scores are 14 to 16 and 12 to 15. Given the high rate of false positives with these recommended cut-off scores, one must ask whether this instrument and its interpretation can ever *really* be so accurate. If it *does* have such accuracy, it would have claims to be a diagnostic tool, not just a screening tool. If, as seems far more likely, it has nothing like such accuracy, the fact that it is presented as having highly accurate cut-off scores leads one to suspect that the SDQ is being credited with the authority of precise quantitative science, far beyond anything which can be supported by scrutiny of its merits. For claims of accuracy and discriminating ability, we seem to be hoodwinked into taking this as “solid science”, claims which should be rejected.

Questions also arise about hidden motives in use of the SDQ. It is almost certainly true that patterns of behaviour are acquired, often at an early age, and once a pattern has started to develop, for good or bad, it is strengthened either by intrinsic reinforcement, or by circular interactions (for instance between parent and child, or in peer groups within which a child’s behaviour is best accepted). One may therefore ask why the intentions of the SDQ have been concealed in the *Practitioner’s Handbook*? There are several possible answers: (a) The agencies/persons who conceived introduction of the SDQ into the *B4 School Check* did not trust the public to accept these intentions; (b) They were uncertain of the merits and potential uses of the SDQ until it had been trialled, and did not want to declare their intentions at an early stage. The correct procedure would, of course, have been to declare it as a research study, and then trial it first on a small-scale; (c) They were not certain that effective intervention was possible or deliverable in practice; (d) They were inexperienced in research methodology.

Serious shortcomings of the SDQ are now recognised. A recent review concluded that “the hypothesised subscales may not always tap distinct constructs”. In addition it was stated “that there are advantages to using the broader internalising and externalising SDQ subscales (one of the models of SDQ data derived by factor analysis) for analyses in low-risk samples, while retaining all five subscales when screening for disorder . . . including those with mental disorder and/or with high

scores on the SDQ subscales”<sup>86</sup>. A further caveat comes in a Dutch study<sup>87</sup> of children in child welfare care (a group with above average mental health problems, often poorly addressed). Compared to their “total difficulties” score, the “impact” supplement to the SDQ was a better predictor of service use, and parents’ or carer’s requests for additional help. The authors state that “caution is warranted when the SDQ is the only source of information for referrals to specialised care”.

(iv) *Policy implications: Lack of Experience in Scientific Methodology.* In the previous subsection, criticism was implied of projects originating in the Ministry of Social Development, yet purporting to deal with health, and using methods of health sciences. Spurious claims to accuracy and the discriminating power of the SDQ betray ignorance of methodology in health sciences. This is not the only example. In a document sponsored by the same Ministry entitled “Vulnerable children”<sup>88</sup>, the subtitle tells all: “Can administrative data be used to identify children at risk of adverse outcomes?” In the “Background” section, we read: “One concern is that operator driven tools rely on the social worker or frontline agency correctly applying the model. Compliance is dependent upon an agent who is sufficiently trained and motivated to apply the model, and to then respond to the estimated risk. A second concern is that operator driven risk assessment tools are infrequently validated for the population being risk rated.” The authors develop an algorithm for predicting childhood maltreatment, using administrative data such as which parents claim a state benefit (plus a variety of other data), in correlation with parallel evidence on maltreatment of children in the same cohort of families. They also compute the proportion of true and false positives, grouped by decile according to “risk” computed with this algorithm. In decile ten, where the risk was estimated to be highest, of all “positive identifications”, 52% were false ones, in decile nine, 63%, and in decile eight, 69.9%. The economics of “intervening” in these high-risk deciles was also computed.

Admittedly, this system is at a research stage, nowhere near implementation, although another document on Vulnerable Children<sup>89</sup>, which *was* to be implemented, cited this model and the algorithm it developed in support. Major

---

86 Goodman A, Lamping,DL and Ploubidis, GB (2010) When to use broader internalizing and externalizing subscales instead of the hypothesized five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. *Journal of Abnormal Child Psychology*, 38, 1179-1191.

87 Janssens,A, Deboutte,D. (2009) Screening for psychopathology in child welfare: the Strengths and Difficulties Questionnaire (SDQ) compared with the Achenbach System of Empirically Based Assessment (ASEBA). *European Child & Adolescent Psychiatry*, 18, 691-700.

88 Vaithianathan,R., Maloney,T, Jiang,N, De Haan,I, Dale,C, Putnam-Hornstein,E, Dare,T. (2012) *Vulnerable Children – Can administrative data be used to identify children at risk of adverse outcomes?* Centre for Applied Research in Economics (CARE), Department of Economics, University of Auckland, September 2012.

<http://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/vulnerable-children/auckland-university-can-administrative-data-be-used-to-identify-children-at-risk-of-adverse-outcome.pdf>

89 Ministry of Social Development (2012) *The White Paper for Vulnerable Children, volume II, Chapter 4.*

ethical issues would need to be addressed before implementation was considered<sup>90</sup>. Nevertheless, questions are raised, more about habits of thought behind such work, than about possibilities in practice: What proportion of false positives is acceptable for intervention at a population level, when intervention might have negative as well as positive impact on individuals in target groups? What is the point of having experts who work with individuals (in this case social workers) if administrators, remote from the ‘coal face’ think that the data to which *they* have access enable superior decisions about the same individuals? Are their predictions *really* superior to those of workers at the coal face? Is it not time to admit frankly ‘We cannot at present predict this’, rather than go through the motions of a pseudo-scientific pretence at prediction?<sup>91</sup> In this context, the following comments from the government’s chief science advisor, Sir Peter Gluckman are highly pertinent:

“Government agencies’ attitudes towards using scientific evidence in making policy is mixed, and at times cavalier, and policies need far more rigorous and objective assessment. . . It’s a worry that some officials have a limited understanding of evidence-based information and do not always apply it, which can lead to poor policy decisions. . . Policy makers can be influenced too much by what they believe or what government ministers might want to hear.”<sup>92</sup>

Let us take an overview of this style of social policy formation informed by the broad sweep of history. The birth of the natural sciences in the seventeenth century depended on reasoning which was mathematical and quantitative, and on strongly-validated concepts with proven ability to support precise reasoning. In the eighteenth century (so-called “Age of Reason”), a version of such science started to be applied in social fields, when the first statistics on age-related life expectancy became available. Statistics seemed to fit neat equations. Not only was this the start of the Life Insurance Industry; in a broader scheme of things, it was the start of what we would now call “social science” and “science-based social policy formation”. The flaw was that, while it used forms of reasoning, and was avowedly quantitative,

---

90 Even when the proportion of false positives and false negatives are much higher than in this study, it may be problematic whether they are acceptable for a screening program. In a Greek study, data collected on hospital admission, supplemented by information from social workers was used to predict child abuse, reaching the impressive figures of 87% sensitivity and 96% specificity. Nevertheless, since the abusing families made up less than 0.1% of all families, the total number of false positives, where there was no abuse, vastly outnumbered the true positive identifications where there was abuse. (Agathonos-Georgopoulos,H., Browne,KD.. 1997, The prediction of child maltreatment in Greek families. *Child Abuse Neglect*, 21, 721-735)

91 A similar issue was raised recently in Britain when government tried to foist the term “dangerous and severe personality disorder” as a diagnosis, but one with neither legal nor medical basis, to be used as a basis for detention of people who had committed no crime. In the end, when community voices and the psychiatric profession combined to express their opposition, this was stopped. One argument coming from psychiatrists was that they were being asked to make predictions which were not possible to make.

92 Radio New Zealand interview, Checkpoint, 9.50 p.m., 3.09.2013: “*Science advisers needed in key Govt departments –Sir Peter.*”

it never established the validity of concepts around which reasoning and data collection were to be based. Over the centuries, another element crept. More and more attention was paid to correlations between different data sets, initially as correlates of suicide statistics, but extending far beyond this, as more facets of social life were charted statistically. With this came increasingly sophisticated methods of statistical analysis. Although causation (which is difficult to define) is different from correlation, the abundance of data on correlations in the academic and policy literature gradually persuaded many (often implicitly, despite explicit denial) to assume that correlation indicated causation. The history of this insidious shift has been described in a recent monograph<sup>93</sup>. In academia, the flaw is usually recognised: Discussion sections of research papers explore alternative ways in which an observed correlation can come about, short of direct causation. However, the shift led to an increasingly doctrinaire empiricism, rather than the balance between empiricism and rationalism which defines the natural sciences. Today, we tend to abandon reasoning, and are in what can be called “The Age of Evidence”.

In areas of social and economic policy, where persuading a naïve populace is the agenda as much as presenting a fair and balanced view, this flaw has been adopted most widely, and most dangerously. Amongst abuses of statistical reasoning is the practice of combining the predictive powers of a number of variables which severally correlate with an outcome, to produce stronger predictions, supposedly even to help identify individuals at risk. So, we are now familiar with public strategies based on “profiling” to screen for possible terrorists in aircraft passenger lists. A similar style prevails in the project already mentioned, at Auckland University, developing an algorithm to predict risk of child maltreatment, on the basis of multiple risk factors. Such statistical methods do have their uses, despite unanswered questions about how risk factors combine – linearly, non-linearly, additively, supra- or sub-additively, or even for one variable to cancel out the effects of another. These matters may be of minor importance when considering population effects, but in areas of social policy, may have decisive impact on individuals. Many papers avoid presenting data as numbers or proportions of people in a cohort, but as group averages on corresponding rating scales, so avoiding confronting the personal impact of a policy. Quite apart from *B4 School*, and the SDQ, the social policy style based on “profiling” of people and on combining risk factors is certainly being used to narrow down those who are potential criminals<sup>94</sup>. Overall, the research style in this area seems to be an inappropriate attempt to mimic methodology of the hard sciences, not least (in view of the extension of the research to many impoverished countries), to give it the aura of universal validity. This is quite properly the strength of these hard sciences, but seldom applies in the social field.

**[VII] For Comparison with SDQ: Merits of Screening Methods to Detect or Identify Problems, in Other Sections of B4 School, and in Related Fields.** In this section, the examples are all related to psychiatry (mainly child psychiatry), or are from other parts of the *B4 School Check*.

---

93 Hacking, I (1990). *The Taming of Chance*. Cambridge University Press.

94 See the following two website addresses: <http://www.opendemocracy.net/ourkingdom/sean-creaney/predicting-criminality-youth-justice-and-early-intervention>;  
<http://gizmodo.com/5517231/crime-prediction-software-is-here-and-its-a-very-bad-idea>

(i) *Continuities between childhood and adult mental disorders.* A review in 1995<sup>95</sup> discussed evidence available at the time on the extent to which adult mental disorders were preceded by related disorders in childhood. Diagnoses included:- depressive disorders (where there are striking continuities between childhood and adult problems, but also complex differences depending on whether a childhood depressive syndrome is or is not linked to conduct disorder); anxiety disorders; anorexia nervosa; conduct disorders; hyperkinetic disorders; autism/pervasive developmental disorders; and specific developmental disorders. The review did not attempt to quantify risk of adult disorder in terms of childhood diagnoses, and much information has accrued since its publication. This includes studies based on long-term follow-up from childhood to adulthood<sup>96</sup>, retrospective studies based on charts<sup>97</sup>, and a review emphasising possible mechanisms for such continuity<sup>98</sup>. The extent of continuity or change of diagnosis is another major issue in such studies.

Childhood forerunners of anxiety disorders deserve detailed coverage. One study, part of a WHO study of childhood mental health in Mauritius<sup>99</sup>, explored the predictive value of “inhibited/uninhibited temperament” in 3 to 11-year-olds. This factor is associated with autonomic signs of arousal: Children who are behaviourally inhibited have higher heart rates, and higher arousal levels as indicated by skin conductance<sup>100</sup>. A behavioural inhibition factor is stable over the years of childhood, a significant fact, since it is a risk factor for anxiety disorders<sup>101</sup>. Further research on this factor has given data on its value in screening for risk of anxiety disorders. One study<sup>102</sup> found this trait, detected in infants aged 21 months, had considerable predictive power for one type of anxiety disorder (“Overanxious Disorder”) at age

---

95 Rutter,M (1995) Relationship between mental disorders in childhood and adulthood. *Acta Psychiatrica Scandinavica*, 91, 73-85.

96 Thomsen,PH (1996) A 22- to 25-year follow-up study of former child psychiatric patients: a register-based investigation of the course of psychiatric disorder and mortality in 546 Danish child psychiatric patients. *Acta Psychiatrica Scandinavica*, 94, 397-403.

97 Dyrborg,J, Wang, A-R. (1997) Adolescent psychiatric patients with child psychiatric debut. A description of diagnostic stability and change. *Nordic Journal of Psychiatry*, 51,153-158.

98 Rutter,M, Kim-Cohen,J, and Maughan,B. (2006) Continuities and discontinuities in psychopathology between childhood and adult life. *Journal of Child Psychology and Psychiatry*, 47, 276-295.

99 Scarpa, A, Raine,A, Venables, PH, Mednick,SA (1995) The stability of inhibited/uninhibited temperament from ages 3 to 11 years in Mauritanian Children. *Journal of Abnormal Child Psychology*, 23, 607-618.

100 Scarpa,A, Raine,A, Venables,PH, Mednick,SA.(1997) Heart rate and skin conductance in behaviourally inhibited Mauritanian children. *Journal of Abnormal Psychology*, 106, 182-190.

101 Biederman,J, Rosenbaum,JF, Hirshfeld,DR, Faraon,SV, Bolduc,EA, Gersten,M, Meminger,SR, Kagan,J, Seidman,N, Reznick,S. (1990) Psychiatric correlates of behavioural inhibition in young children of parents with and without psychiatric disorders. *Archives of General Psychiatry*, 47, 21-26; Hirshfeld,DR. Rosenbaum,JF, Biederman,J, Bolduc,EA, Faraone,SV, Snidman,N, Reznick,JS, Kagan,J. (1992) Stable behavioural inhibition and its association with anxiety disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 31,103-111; Biederman,J, Rosenbaum,JF, Bolduc-Murphy,EA, Faraone,SV, Chaloff,J, Hirshfeld,DR, Kagan,J. (1993) A 3-year follow-up of children with an without behavioural inhibition. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32,814-821.

102 Biederman,J, Rosenbaum,JF, Hirshfeld,DR, Faraon,SV, Bolduc,EA, Gersten,M, Meminger,SR, Kagan,J, Seidman,N, Reznick,S. (1990) Psychiatric correlates of behavioural inhibition in young children of parents with and without psychiatric disorders. *Archives of General Psychiatry*, 47, 21-26.

7-8 years: Amongst a group of 49 children, this predictor, with the cut-offs used, had 0% false positives, and 72% false negatives (or sensitivity, 100%; specificity, 28%). Use of higher threshold levels would have given increased specificity, though at cost of decreased sensitivity. Other diagnoses, especially other anxiety disorders, were also found at elevated rates, but (except for Oppositional Defiant Disorder), not to a statistically significant degree. A second study<sup>103</sup> of infants assessed at age 1 year, and again at age 13, gave data using two cut-off thresholds for Social Anxiety Disorder. At a lower threshold for social anxiety in adolescents there were 22% false positives and 11% false negatives. At a higher threshold, at which social anxiety in adolescents led to definite impairment in everyday life, in females, there were 31.7% false positives, and 2.4% false negatives. For males the values were 42% and 6.1%. Combining genders, the values were 36.5% and 4.1%. A third study<sup>104</sup> found that behavioural inhibition, documented in infants aged 14-24 months, predicted adolescent Social Anxiety Disorder at age 14-16 years with only 13.1% (16/122) false positives, and 9.8% (12/122) false negatives. Clearly, in these controlled studies, the behavioural inhibition index for infants and children, is a powerful predictor of some forms of anxiety disorder in adolescence, a far better predictor, in fact than the SDQ for Conduct Disorder or any other diagnosis. However, both studies were of highly selected small groups of children. It is far from certain that this predictive power would apply in a population-based study, with acceptably-low levels of false positive and false negative determinations.

(ii) *Tests used in B4 School, other than SDQ:* In auditory tests for school-age children, similar to those used in *B4 School*, a study from Texas<sup>105</sup> found very low false positive rates (1.2%, 4.2% and 6.4% for different components). Different tests had different sensitivity and any one did not always pick up abnormality. To avoid false negatives it was recommended that the three be combined. Visual screening of pre-schoolers, achieves similar high sensitivity (92%) and specificity (97%)<sup>106</sup>.

In contrast to these two reliable, specific and sensitive testing systems, the other behavioural component of *B4 School*, the PEDS, performs far less well. It has been found to have sensitivity and specificity ranging from 70% to 80%, across all ages for which it is used<sup>107</sup>. This is considerably inferior to auditory and visual screening in *B4 School*, but substantially better than the SDQ in most studies.

(iii) *Other screening methods for child mental health problems.* We are dealing here with mass screening programs, and the objective is to screen a large number of

---

103 Schwartz,CE, Snidman,N. Kagan,J. (1999) Adolescent social anxiety as an outcome of inhibited temperament in childhood. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38,1008-1015.

104 Chronis-Tuscano,A, Degnan,KA, Pine, DS, Perez-Edgar,K, Henderson,HA, Diaz, Y, Raggi,VL, Fox, NA (2009) Stable early maternal report of behavioural inhibition predicts lifetime social anxiety disorder in adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 928-935.

105 Krueger, WWO, Fergusson,L (2002) A Comparison of Screening methods in School-Aged children. *Otolaryngology – Head and Neck Surgery*, 127, 516-519.

106 Lennerstrand,G, Kvarnstrom,G, Jakobsson,P. (1998) Screening for visual and ocular disorders in children, evaluation of the system in Sweden. *Acta Paediatrica*, 87, 1173-1170.

107 Smith, PK, Schor,E. (2005) Enhancing Child Development Services in Medicaid Managed Care. Center for Health Care Strategies, Inc, Hamilton, New Jersey. ([www.chcs.org/usr\\_doc/Toolkit.pdf](http://www.chcs.org/usr_doc/Toolkit.pdf))



people. A primary concern is then not so much the accuracy in itself but rather efficiency in first-approximation assessment. A difficulty arises from the fact that diagnoses in childhood psychiatry are even less secure than in adults. In computing sensitivity and selectivity of any instrument, it is then difficult to find a “Gold Standard” (a reliable diagnosis) for comparison. As a result comparisons are based on a variety of methods, none completely satisfactory. The best standard is a clinical interview and diagnoses this leads to, but other methods include comparison with previous population-based prevalence statistics, final outcome in a longitudinal study, or comparison with other rating scales (however poorly validated they may be). Some studies screen for a broad range of child mental health problems, others aim at specific diagnoses. Even when screening is for a broad range of diagnoses, published reports usually give only mean values averaged across diagnoses for selectivity and sensitivity. The fact that individual diagnoses are poorly identified is then not made clear.

Using screening methods evaluated in relation to a clinical interview, it has been possible to achieve useful levels of sensitivity and selectivity in non-specific mental health screening of children<sup>108</sup>, in assessments related to the need for an individual special education program<sup>109</sup>, or for identifying the broad classes of internalising and externalizing problems<sup>110</sup>. (There is usually difficulty in achieving one of these requirements without sacrificing the other.) The most successful screening based on clinical interviews as the Gold Standard used The Brief Child and Family Phone Interview (BCFPI), a 36-item protocol related to DSM-IV diagnoses<sup>111</sup>, based on parental evaluation, using trained lay interviewers. There is no report of how long the phone interview usually took. Based on 399 children and adolescents aged 5-17 years<sup>112</sup>, the following values were given for sensitivity and selectivity:

	<i>Sensitivity</i>	<i>Selectivity</i>
ADHD	69.9	73.1
ODD	77.4	71.4
CD	56.3	89.3
Ext	81.6	62.7
SAD	49.1	89.9
GAD	39.2	93.1
MDD	37.7	89.4
Int	58.3	79.9

---

108 Navon, M, Nelson,D, Pagano,M, Murphy,M. (2001) Use of the Pediatric Symptom Checklist in strategies to improve preventive behavioural health care. *Psychiatric Services*, 52, 800-804.

109 Barbarin, OA (2007) Mental Health Screening of Preschool Children: validity and Reliability. *American Journal of Orthopsychiatry*, 77, 402-418.

110 Brasil,HHA, Bordin,IA. (2010) Convergent validity of K-SADS-PL by comparison with CBCL in a Portuguese speaking outpatient population. *BMC Psychiatry*, 10, 83.

111 Cunningham,CE, Boyle,MH, Hong,S, Pettingill,P, Bohaychuk,D. (2009) The Brief Child and family Interview (BCFPI): 1: Rationale, development, and description of a computerized children’s mental health intake and outcome assessment tool. *Journal of Child Psychology and Psychiatry*, 50, 416-423.

112 Boyle,MH, Cunningham,CE, Georgiades,K, Cullen,J, Racine,Y, Pettigill,P. (2009) The Brief Child and family Phone Interview (BCFPI): 2. Usefulness in screening for child and adolescent psychopathology. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 50, 424-431.

This method appears to obtain a better combination of sensitivity and selectivity for externalising disorders than for internalising ones. Data on prevalence obtained with the BCFPI, suggest it may underestimate numbers in the latter class. However, it is equally good across gender or age groups (5-11 years versus 12-17 years).

Screening methods standardised in other ways are generally not so good. Their weaknesses include their not having been evaluated for a broad range of childhood mental health problems, or having unacceptably-low levels for either sensitivity or selectivity. For methods standardized in relation to other instruments, few studies report values for sensitivity and selectivity, although one paper<sup>113</sup> gives acceptable values for an “any problem” determination, and another<sup>114</sup> reports acceptable values for both of the broad categories of externalising and internalizing disorders. This is helpful, since many methods of screening appear biased towards one or other of these categories (as is the SDQ). The overall conclusion is that there are screening methods available for both detecting and identifying several major classes of child mental health problems which are better than the SDQ. However, more specialised instruments are needed to screen for particular problems (such as autism, obsessive-compulsive disorder/Tourette’s syndrome, PTSD, or problems arising primarily from family dysfunction). These method need to be used only when such disorders are suspected or in high-risk groups, rather than in mass screening programs.

(iv) *Early prediction of schizophrenia and depressive/anxiety disorders:* It is generally recognised that the disorder called schizophrenia is characterised not only by psychotic episodes, but also by on-going psychological or psychophysiological traits. The latter are potential predictors of future illness, for use in screening. Here we are concerned with their possible use as *early* predictors (in infancy or early childhood). One of the first clues that this was possible came from a study of video footage taken by parents, of children who later developed schizophrenia, compared with that of children who did not<sup>115</sup>. Analysis of facial expressions showed clear differences between groups in emotional responsiveness. For expressions of joy a deficit was seen in females destined to get the diagnosis as early as 4-8 mo ( $p=0.03$ ; 1-4y:  $p=0.003$ ), as well as at later stages of childhood (10-13y  $p=0.04$ ; 13-16y,  $p=0.004$ ). In males early differences were also seen, but developed in more complex ways (1-4 mo:  $p=0.04$ ; but trend [ $p=0.06$ ] toward *more* joy at 13-16y). Of course, facial indications of joy are imprecise markers: Reduced joy may point to a general inability to cope with a specific environment for any number of reasons, which may be related to cognition, perception, or general bodily well-being (none of which are primary abnormalities of emotion). More precise markers were suggested in work based on the British National Child Development Study<sup>116</sup>, comparing data obtained

---

113 Kresanov,K, Tuominen,J, Piha,J, Almqvist,F. (1998) Validity of child psychiatric screening methods. *European Child & Adolescent Psychiatry*, 7, 85-95.

114 Cameron, FJ, Smidts,D, Hesketh,K, Wake,M, Northam,EA (2003) Early detection of emotional and behavioural problems in children with diabetes: the validity of the Child Health Questionnaire as a screening instrument. *Diabetic Medicine*, 20, 646-650.

115 Walker,EF, Grimes,KE, Davis,DM, Smith,AJ (1993) Childhood precursors of schizophrenia: Facial expressions of emotion. *American Journal of Psychiatry*, 150, 1654-1660.

116 Done,DJ, Crow,TJ, Johnstone,EC, Sacker,A. (1994) Childhood antecedents of schizophrenia and affective illness: social adjustment at ages 7 and 11. *BMJ*, 309, 699-703.

at ages 7 and 11 years with that on adult admissions to psychiatric hospitals. At age 7, children who later developed schizophrenia were rated by teachers as socially less-well adjusted than controls, a difference which was more apparent in boys than girls, and related to over- rather than under-reactive behaviour. By the age of 11 children who later obtained diagnoses in the neurosis domain, particularly girls, had increased maladjustment (including over-reactions and under-reactions).

Several studies sought predictors of schizophrenia in early teenage years rather than childhood or infancy, based on pre-psychotic, or prodromal symptoms, or attenuated versions of psychotic symptoms<sup>117</sup>. Others examine early predictors, but report group-averaged data, rather than the numbers or proportions of individuals detected by the potential screening index<sup>118</sup>. Such studies do not serve our purposes of providing evidence of a viable screening index, although they identify possible markers for developing schizophrenia. Emotional problems (internalising problems) may be one such, but not very specific since these are also markers for developing anxiety problems, although, one study, found them to be more severe in those who later receive the schizophrenia diagnosis<sup>119</sup>. More specific markers in this study were significant impairments in neuromotor function, receptive language, and cognitive development. Such impairments were present in addition to emotional problems among children later receiving the diagnosis of schizophreniform disorder. Similar findings have been made for the full schizophrenia diagnosis<sup>120</sup>, including evidence from another study based on childhood videotapes<sup>121</sup>.

Amongst these papers is a single one<sup>122</sup> which gives data allowing one to assess the merits of such markers as very early screening indices. Using data collected at child psychiatry clinics at the Maudsley hospital, abnormal suspiciousness or sensitivity, and relationship difficulties with peers were found to be associated with later schizophrenia. In contrast, affective psychosis was associated with childhood “hysterical” symptoms (depersonalization, conversion hysterical symptoms, non-epileptic disturbance of consciousness, any of which by themselves were rare), gross overactivity, morbid depression, hypochondria and disturbances in eating. For schizophrenia, positive sensitivity was 81.2%, while the specificity taking all data

---

117 Poulton,R, Caspi,A, Moffitt,TE, Cannon,M, Murray, RM, Harrington, H. (2000) Children’s self-reported psychotic symptoms and adult schizophreniform disorder. *Archives of General Psychiatry*, 57, 1053-1058.

118 Johnstone,EC, Ebmeier,KP, Miller,P, Owens,DGC, Lawrie,SM. (2005) Predicting schizophrenia: findings from the Edinburgh high-risk study. *British Journal of Psychiatry*, 186, 18-25.

119 Cannon,M, Caspi,A, Moffitt,TE, Harrington,H, Taylor,A, Murray,RM, Poulton,R. (2002) Evidence for early-childhood pan-developmental impairment specific to schizophreniform disorder: Results from a longitudinal birth cohort. *Archives of General Psychiatry*, 59, 449-456.

120 Isohanni, M, Murray, GK, Jokelainen,J, Croudace,T, Jones, PB. (2004) The persistence of developmental markers in childhood and adolescence and risk for schizophrenic psychoses in adult life. A 34-year follow-up of the Northern Finland 1966 birth cohort. *Schizophrenia Research*, 71, 213-225.

121 Schiffman,J, Walker,E, Morten,E, Schulsinger,F, Sorensen,H, Mednick,S. (2004) Childhood videotaped social and neuromotor precursors of schizophrenia: a prospective investigation. *American Journal of Psychiatry*, 161, 2021-2027.

122 Cannon,M, Hollis,C, Kargin,M, Taylor,E, Murray,RM, Jones, PB (2001) Predictors of later schizophrenia and affective psychosis among attendees at a child psychiatry department. *British Journal of Psychiatry*, 178, 420-426.

was 65% (false positive rate of 35%). For affective psychoses, 7/11 cases were correctly classified (36.4% false negatives), while the false positive rate was high (56.2%). (The poor predictive performance for affective psychosis was partly due to loss of data loss about hysterical symptoms).

A recent population-based birth cohort study<sup>123</sup> sought predictors of major depression anxiety, and overall mental health problems at age 21, using CBCL data on internalising and total problems collected at age 5. The data were poor predictors: For depression and anxiety, although specificity was high (77-94%, depending on the cut-off point on the “internalising” score) sensitivity was low (6.7-25.1% for depression; 8.8-28.3% for anxiety). In predicting “any mental health problem”, the “total problems” CBCL score fared no better.

(v) *Predicting delinquency, aggressive behaviour and criminal tendencies from early childhood data.* If we follow the apparent hidden agenda behind use of SDQ in *B4 School*, and the implication, that, by using it, one might be able to intervene to prevent delinquency and crime in adolescence and adulthood, another question arises: *Can childhood markers of behavioural problems predict antisocial patterns of behaviour in later life?*<sup>124</sup>

Before we consider evidence, we need to clarify some conceptual issues. A review written in 1987<sup>125</sup> discussed two major classes of externalizing behaviours in childhood - conduct disorder and hyperactivity/inattention. The two are seen as closely related, but the author argues that important distinctions should be made. At the time, instruments used to rate these problems were confounded, because of the assumption that the two were related, and because studies of the topic used cohorts which were a haphazard mix of the two problems. At the time of writing, some studies suggested the merits of subdividing the concept, although syndromal independence of the subdivisions had not been established. Several methods can be used to clarify such issues. Factor analysis is a statistical method for establishing the unity of – or subdivisions within – complex data sets of psychiatric symptoms. Several factor analyses have been conducted on children aged 10 years or less on SDQ data<sup>126</sup>. Another approach is to seek differential correlations of symptoms

---

123 Najman, JM, Heron,MA, Hayatbakhsh,MR, Dingle,K, Jamrozik,K, Bor,W, O’Callaghan,MJ, Williams,GM. (2008) Screening in early childhood for risk of later mental health problems: a longitudinal study. *Journal of Psychiatric Research*, 42, 694-700; see also: Dingle,K, Alati,R, Williams,GM, Namjan,JK, Bor,W, Clavariono,A. (2009) The ability of YSR DSM-oriented depression scales to predict DSM-IV depression in young adults: a longitudinal study. *Journal of Affective Disorder*, 121, 45-51.

124 This topic was featured in a recent article in *Observer* (12.05.2013): Tim Adams: *How to spot a murderer’s brain.*

125 Hinshaw, SP (1987) On the distinction between attentional deficits/hyperactivity and conduct problems/aggression in child psychopathology. *Psychological Bulletin*, 101, 443-463.

126 Smedje,H, Broman, J-E, Hetta,A-L, Knorrning,A-L. (1999) Psychometric properties of a Swedish version of the “Strengths and Difficulties Questionnaire”. *European Child & Adolescent Psychiatry*, 8, 63-70; Hawes,DJ, and Dadds,MR (2004) Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Australian and New Zealand Journal of Psychiatry*, 38,644-651; Kashala,E, Elgen,I, Sommerfelt, K, Tylleskar, T. (2005) Teacher ratings of mental health among school children in Kinshasa, Democratic Republic of Congo. *European Child & Adolescent Psychiatry*, 14, 208-215; Van Leeuwen,K, Meerschart,T, Bosmans,G, De Medts,L, Braet,C. (2006) The Strengths and Difficulties Questionnaire in a Community Sample of Yung

clusters with other variables.<sup>127</sup> A third approach is to seek differentials in the longitudinal course of disorders characterised by different syndromes<sup>128</sup>. Altogether, these studies show that hyperactivity/inattentiveness is to some degree separate from conduct disorder, although the separation is not sharp. Further lack of clarity arises because conduct disorder is itself not uniform, but may be resolvable into separate components (oppositional versus aggressive), and because psychological traits seen in one environment may not transfer to another.

Turning to the substance of this section, the possibility of screening to detect those at risk of delinquency, aggressive behaviour or criminal tendency, evidence about risk factors is reviewed. These can be subdivided into biological risk factors acting on the brain, and traits expressed in detectable psychological measures. Birth complications are an early influence on brain biology. In a study from Copenhagen<sup>129</sup> on predictors of violent crime by age of 18 years in a cohort of 4269, this was found to have no effect by itself, but when combined with maternal rejection had a major effect. While only 4.5% of the adolescents had both risk factors, they account for 18% of all violent crimes, an effect specific to such crimes. Of course, it is possible that covariates of each risk factor were more directly related to the association than the two variables studied. A study from the same researchers<sup>130</sup>, with similar design collected data from a randomised subset of the original large cohort. Data included those on obstetric complications, non-maturity at birth, neurological status over the first five day post-natal period, motor development at 1 year, family, social, demographic, and behavioural measures at age 17 to 19 years, and criminal data at 20 to 22 years of age. Participants were divided into three groups, those with obstetric and neuromotor risk factors, those in the poverty class and those with both these risk factors. Rates of crime were increased in all three groups compared with a control group with neither risk factor; and the increase in those with combination of risk factors was approximately double that in the other two groups with single risk factors. The authors conclude: "Because this biosocial group accounted for 70.2% of all crimes committed in the entire sample, early interventions that tackle these deficits might significantly reduce violence." There is a fundamental flaw in this conclusion: The control group

---

Children in Flanders. *European Journal of Psychological Assessment*, 22, 189-197; Sanne,B, Torsheim,T, Heiervang,E, Stormack,KM (2009) The Strengths and Difficulties Questionnaire in the Bergen Child Study: A Conceptually and methodically Motivated Structural Analysis. *Psychological Assessment*, 21, 352-364.

127 Stormshak,EA, Bierman,KL (1998) The implication of different developmental patterns of disruptive behavior problems for school adjustment. Conduct problems Prevention Research Group. *Development & Psychopathology*, 10, 451-467.

128 Jester, JM, Nigg, JT, Adams, K, Fitzgerald, HE, Puttler, LI, Wong, MM, Zucker, RA (2005) Attention/hyperactivity and aggression from early childhood to adolescence: Heterogeneity of trajectories and differential influence of family environment characteristics. *Development & Psychopathology*, 17, 99-125.

129 Raine, A, Brennan, P, Mednick, SA (1994) Birth complications combined with early maternal rejection at age 1 year predispose to violent crime at age 18 years. *Archives of General Psychiatry*, 51, 984-988.

130 Raine, A., (1996) High rates of violence, crime, academic problems, and behavioral problems in males with both early neuromotor deficits and unstable family environments. *Archives of General Psychiatry*, 53, 544-549.

of 44 persons was not part of the original cohort, and made up only 10% (44/441) of the total number from which this 70.2% figure was derived. It is beyond belief, that, in a representative sample, only 10% would have neither of the specified risks. Therefore, the 70.2% figure is a massive over-estimation of the predictive value of the variable studied. To jump from that figure to talk of early intervention, presumably on a population-wide basis, is a serious, and potentially dangerous *non sequitur*.

A third paper from the same group<sup>131</sup> uses data collected during a long-term study of child mental health in the Indian Ocean island of Mauritius (described earlier<sup>132</sup>). From an original cohort of 1800 randomly selected from the whole population, resting heart rate data at age 3 years were available for 1,795 male and female children. Evidence was found linking low resting heart rate in infants specifically to aggressive forms of antisocial behaviour at age 11 years. However, low and high heart rate groups made up only 20.2% of the total sample, that is those whose heart rates were 1SD above or below the mean for the whole group. In this fraction, figures given indicate that the false positive rate (with a high heart rate, but subsequently found to be aggressive) made up 33.5% of the deemed-non-aggressive group, while the false negative rate (low heart rate, later found to be non-aggressive) made up 35.5% of the deemed-aggressive group. The authors claim to have found “an important diagnostically specific, well-replicated, early biological marker for later aggressive behaviour”. Several covariates were found to reduce the strength of the association (though not to eliminate it). However, the predictive value is limited by the fact that only 20% of the original cohort was used in the analysis. The authors make no claim that heart rate would be a viable mass screening measure to predict aggressiveness, but the claims they do make seem excessive.

Apart from such clearly biological predictors, other papers examine using psychological or behavioural trait markers, documented in infants or children as potential predictors. Such markers may themselves be affected either by underlying brain biology, or by life experiences and social influences. The possible role of maternal rejection in one of the Copenhagen papers was already discussed. Another study<sup>133</sup> documents the role of family cohesiveness and stimulating home environments as protection against emergence of hyperactivity/inattention and/or aggressiveness between school entry and age 18 years. Family cohesiveness protected against aggressiveness, but not against attention/hyperactivity. Early stimulation protected against development of attention/hyperactivity but not against aggressiveness. This study provided no data on sensitivity or selectivity of the

---

131 Raine, A, Venables,PH, Mednick,SA (1997) Low resting heart rate at age 3 years predisposes to aggression at age 11 years: evidence from the Mauritius Child Health Project. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 1457-1464.

132 Schulsinger,F, Mednick,SA, Venables,PH, Raman,AC, Bell,B. (1975) Early detection and prevention of mental illness: The Mauritius project. A preliminary report. *Neuropsychobiology*,1, 166-179.

133 Jester, JM, Nigg, JT, Adams, K, Fitzgerald, HE, Puttler, LI, Wong, MM, Zucker, RA (2005) Attention/hyperactivity and aggression from early childhood to adolescence: Heterogeneity of trajectories and differential influence of family environment characteristics. *Development & Psychopathology*, 17, 99-125.

predictors in relation to outcome, so the merits of the predictor variable as screening tools at the individual level is unclear.

A fourth study, started in 1991-1993 in Worcester, Massachusetts<sup>134</sup>, explored the value in predicting adverse educational and behavioural outcomes of two measures made in pre-schoolers - disruptive behaviour, and what was called “adaptive disability” (a significant discrepancy between intelligence and daily adaptive functioning). This was assessed in pre-schoolers, of whom 43 had neither early risk factor, 98 had disruptive behaviour only, and a further 28 had both disruptive behaviour and adaptive disability. As percentages of each group showing various adverse outcomes after three years, there certainly was a highly significant increase, starting with those deemed not at risk, to those with one risk factor, to those with both risk factors. However, although these results identify a risk factor, when the data presented are recalculated as numbers rather than percentages with adverse outcomes in each group, and then to compute percentages of false positives and false negatives, it becomes clear that this complex measure is no use at all as a screening tool (see Table III). In addition a high proportion of those parents invited to take part on behalf of their children, declined the invitation, and the control group was probably a massive numerical underestimate of those in the general population with neither risk factor. Data presented are thus worthless, as pointers to a viable screening tool for use in early intervention programs (which parents had been told was the objective of the study).

**TABLE III**  
**Adaptive Disability as a Predictor of Disruptive Behaviour**

	Control n=43	DB only n=98	DB+AD n=28	False +ves		False -ves	
				DB	DB+AD	DB	DB+AD
School Conference	3	33	15	2.1%	4.2%	76.6%	78.9%
Grade retention	1	3	1	0.7%	1.4%	97.9%	98.6%
Suspended from school	0	1	10	0%	0%	99.3%	85.9%
Special education	6	22	13	4.2%	8.4%	84.4%	81.7%
Child in therapy	1	18	8	0.7%	1.4%	87.2%	88.7%
Child on psych meds	0	14	8	0%	0%	87.2%	88.7%

A fourth study was based in four cities in the USA<sup>135</sup>, as part of the Conduct Problems Prevention Research Group, started in 1992 (see below). At each of four

134 Barkley,RA, Shelton,TL, Crosswait,C, Moorehouse,M, Fletcher,K, Barrett,S, Jenkins,L, Metevia,L. (2002) Preschool children with disruptive behavior: Three-year outcome as a function of adaptive disability. *Development & Psychopathology*, 14, 45-67.

135 Flanagan,KS, Bierman,KL, Kam,CM (2003) Identifying at-risk children at school entry: The usefulness of multibehavioral problem profiles. *Journal of Clinical Child & Adolescent Psychiatry*, 32, 396-407.

sites 100 children in their first school year were randomly selected as the normative group, to be combined with a similar number selected as the “at-risk” group, making a total number of participants of 755. This selection was based on teacher ratings made of children in their first year, on aggressiveness, hyperactive-inattentiveness, and deficits in pro-social behaviour. These data were used as predictor variables for outcome two years later, with respect to Behavioural Outcome (aggressive-disruptive behaviour, for which 35% of variance was accounted for), Social Outcome (social preference<sup>136</sup>, for which 19% of variance accounted for), and Academic Performance (18% of variance accounted for). Specificities were generally very high (for Behavioural Outcome: 79-97% using different models), for Academic outcome, 76-95%, and for Social Outcome: 81-97%. Sensitivity was lower (Behavioural Outcome 20-80%; Academic outcome: 7-63%; Social outcome: 14-74%). Specificity looks very impressive, implying false positive rates around 10%. However, there is a major flaw: The normative group was likely to be under-represented compared to the general population, probably massively under-represented. Therefore false positives in a representative cohort would be under-represented, probably massively.

The study most relevant to use of the SDQ as part of *B4 School*, published in 2005<sup>137</sup>, was part of a 25-year longitudinal study of development in 973 children in the Christchurch region of New Zealand. It examined the extent to which data on child behaviour, collected at age 7-9, could predict (at age 21–25 years) adverse outcomes such as crime, substance dependence, mental health problems, difficulties with sexual or partner relationships, and in education and employment. According to the level of childhood disturbance, participants were divided into four subgroups, with cut-offs at the 50<sup>th</sup>, 30<sup>th</sup>, 15<sup>th</sup> and 5<sup>th</sup> percentile of increasing severity. After control for confounding factors, association between early conduct problems and subsequent education or employment outcomes became statistically non-significant. However, associations persisted for other outcomes (crime, substance dependence, mental health and sexual/partner relationships). Children in the most disturbed 5% had rates of these outcomes between 1.5 and 19 times higher than those for the least disturbed 50%, this applying equally to males and females. Association between predictor and outcome variables is presented as percentages for the various crime categories in the various conduct disorder groups, not as correlation coefficients or as percentage variance accounted for. However, it is clear that even in the most severely disturbed children, the majority did not go on to become criminals. Clearly the vast majority of the variance is unaccounted for by predictor variables. One may ask how useful such results are for screening at age 7-9 (or even earlier, as in *B4 School*). A guide to this comes by multiplying the rate of each type of adult adverse

---

136 Children were asked to identify classmates they liked best [LM] and liked least [LL]. The total number of LM and LL nominations received by each child from his or her peers were calculated, standardised within the classroom, and the difference between the two was used to generate a social preference score. This process raises ethical issues, which is troubling, given that there was no evidence of any ethical clearance in this study.

137 Fergusson DM, Horwood LJ, Ridder EM. (2005) Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood. *Journal of Child Psychology & Psychiatry & Allied Disciplines*. 46, 837-849.



outcome (as a percentage), by the proportion of the cohort to which this applies (Table IV). It is clear that identifying those most at risk and intervening on their behalf would have only a minor effect on later crime statistics.

	percentile	50%	30%	15%	5%
TOTAL	973	486.5	281.9	140.9	48.6
<i>Adverse outcome (Crime)</i>					
Property		35.4%	30.7%	22.6%	11.3%
Violence		32.25%	30.2%	24.4%	13.1%
Arrested/convicted (non-traffic)		30.1%	30.3%	25.6%	14.0%
Repeat (5+) traffic offenses		35.5%	30.8%	22.5%	11.2%
Ever imprisoned		14.6%	25.2%	32.6%	27.0%

How does this paper relate to *B4 School*? This paper is not cited in the *Handbook for Practitioners*, and uses predictor variables on conduct disorder at age 7-9 years, not pre-school ages. However, it is cited in the *Interagency Plan for Conduct Disorder/Severe Antisocial Behaviour*, one of whose authors is also the first author of the paper just discussed. On p. 43 of the *Interagency Plan*, a timetable for action is outlined, as follows:

<b>Phase one, 2007-2008</b>
* Systematic screening and eligibility processes for the education sector, developed, piloted and published.
* Gradually increase the proportion of Ministry of Education behaviour clients in the 3-7 age group receiving evidence-based parenting, child and teacher programmes.
* Begin investment in size, skills and training of Ministry of Education specialist workforce, including development of a group of experienced practitioners accredited in mentor and trainer roles.
* Complete design work on Centre of Excellence, with advice from Experts' Group, and begin implementation.
<b>Phase Two 2009-2010</b>
* Ongoing expansion of parent, teacher and child components of Ministry of Education Services.
* Complete establishment of Centre of Excellence to test and refine the comprehensive behaviour intervention
* Implement systematic screening and assessment processes for all 3-7 year-olds within the education sector
* Continue to develop mentor and trainer role.
<b>Phase Three 2011-2012</b>
* Ministry of Education behaviour clients aged 3-7 years begin a comprehensive programme within 12 months of becoming eligible.

- \* Expand provision to ensure that children requiring a comprehensive behavioural intervention (up to 5 percent of children) receive this level of intervention before they are eight years old
- \* Review the comprehensive intervention offered by the Ministry of Education and consider whether further refinement or additional components are required.
- \* Mentor and trainer group accredited.

The style here smacks of a “chain of command”, quasi-military style of administration between government agency and providers, with little allowance for feedback, the negotiation implicit in the concept of informed consent, or that there be any conceivable barriers or constraints to publication. It is particularly regrettable in the mental health area, where so much delivery of services is based around mutuality and relationships. The main point however, is that the several mentions of 3-7 year-olds in this section of the *Interagency Plan* can only be referring to the soon-to-be-implemented *B4 School*, yet these two documents fail to cite each other.

(vi) *Policy implications: Over-Enthusiasm for Preventive Medicine.* In the last twenty years, the phrase (or “mantra”) “early intervention” has been much used and the concept promoted, as a way to solve a wide variety of health problems. Preventive medicine started with mass vaccination and immunization programs, where the scientific basis was often precise, and dangers of out-of-control epidemics tangible to specialist and lay persons alike. This is the heart-land of preventive medicine; yet even for interventions of that sort, there has been, and often still is, continuing controversy. Preventive medicine is now being applied in areas of much greater complexity, where dangers inherent in amplification of population-level disease due to spread of infection do not apply, where screening is imprecise, and intervention less certain. There is now a tendency to over-enthusiastic and misjudged advocacy for screening and early intervention in areas where their value is uncertain. Alternatively, programmes are implemented energetically at considerable cost to taxpayers, without the balance between merits and disadvantages having been thought through, and without the programme having been brought into the headlights for public debate (as though “Big Brother Knows Best”). A recent paper from the UK criticises over-diagnosis of dementia in the elderly as part of a screening program<sup>138</sup>. Another instance, in New Zealand, is the screening/intervention programme to prevent rheumatic heart disease<sup>139</sup>, about which, we read, from the Ministry of Health: “School-based throat swabbing of primary and intermediate-aged children is a key component of the four-year program, but there are fears that, once swabbing stops, the situation will revert to square one”. “Patients . . . will start to rely on the service instead of seeing their GP.” Anonymous informants, including front-line health professionals, conveyed the message that “more focus was needed on social factors to eradicate the disease”.

In writing this, apparently opposing the principle of preventive medicine, I must also assert that I see the other side of the coin as well. In my own specialist field of research, on the complex disorder called schizophrenia, I am aware of, and mainly

---

138 Le Couteur, DG, Doust, J, Creasey, H, Brayne, C. (2013) Political drive to screen for pre-dementia: not evidence based and ignores the harms of diagnosis. *BMJ*, 347, f5125.

139 Bronwyn Torie: *Rheumatic fever drive set to fail – report* Dominion post, 8.04.2013.

accept the arguments about the long-term damage done by illness which is undetected and untreated in its first few years. I have long sought ways to implement an early intervention program, and have contributed to design of a psychological instrument to screen adolescents to narrow down and alert clinicians to those at possible risk. I am no naïve critic.

In the year 2002, the author of a review on science-based intervention in communities<sup>140</sup> writes as follows: “In the past decade, prevention science has emerged as a discipline built on the integration of life course development research, community epidemiology, and preventive intervention trials. . .Prevention science postulates that negative health outcomes like alcohol abuse and dependence can be prevented by reducing or eliminating risk factors and enhancing protective factors in individuals and their environments during the course of development. Despite the advances in the science base for effective preventive interventions and the investments in community-wide preventive interventions, many communities continue to invest in prevention strategies with limited evidence of effectiveness.”

## **G: Evidence on effectiveness of intervention in childhood years as a public health/social development measure.**

(i) *Introduction.* The aim of this section is to review research literature in studies with objectives similar to those underlying use of the SDQ in *B4 School*. Before going into detail, readers are reminded of National Health Committee Guidelines already mentioned, on health screening programs. There should be “an effective and accessible treatment or intervention for the condition indentified through early intervention”, and “high-quality evidence, ideally from randomised controlled trials [should] shows that a screening programme is effective in reducing morbidity or mortality”.

A generation ago, when the possible screening for, and intervening to prevent conduct disorders started to be discussed<sup>141</sup>, topics such as low referral rate from providers of primary care, excess of false positives, and cost-effectiveness balance were already on the agenda. The earliest studies on the efficacy of intervention to prevent conduct disorders<sup>142</sup> clearly showed potential benefits of such intervention. However, they recruited participants from child health clinics, which, in terms of proving effectiveness, is very different from screening and intervention on a mass scale.

---

140 Hawkins,JD, Catalano,RF, Arthur,MW (2002) Promoting science-based prevention in communities. *Addictive Behaviors*, 27, 951-976.

141 Nilsson, C, Sundelin,C, Vuille JC (1977) General health screening of four-year-olds in a Swedish country. V. A Strategy for improving the effectiveness and the cost efficiency of the psychological screening program. *Acta Paediatrica Scandinavica*, 66, 289-295; Brink, SG, Nader, PR (1984) Comprehensive health screening in elementary schools: an outcome evaluation. *Journal of School Health*, 54, 75-78.

142 Webster-Stratton,C. (1984) Randomized trial of two parent-training programs for families with conduct-disordered children. *Journal of Consulting & Clinical Psychology*, 52, 666-678; Webster-Stratton,C. (1988) Self-administered videotape therapy for families with conduct-problem children: comparison with two cost-effective treatments and a control group. *Journal of Consulting & Clinical Psychology* 56, 558-566.

The Church Report (section 9, part 2) dealing with interventions on parenting for preschoolers, with details of previous programmes and their results, concludes that “parenting skills training in any of the variants described above can have a major impact, helping somewhere between 50% and 60% of the parents of young antisocial children to make changes which are necessary in order to return the young antisocial child to a normal developmental pathway.” Documentation is detailed and scholarly. There is however, a big difference between showing efficacy in principle in a research study, and showing its effectiveness in practice, when rolled out for a whole population. In addition to ensure an good balance between false positives and false negatives, further issues arise when attempting to implement scientific findings in practice. Amongst these are: (i) ensuring that a pilot research study recruits a representative sample of the population for whom the intervention is designed. Without this estimates of false positives and negatives would be inaccurate and could exaggerate the merits of a potential screening programme; (ii) avoiding sampling bias due either to attrition before the final outcome data are collected, or due to exclusion criteria used in selecting cohorts in pilot studies; (iii) in conducting pilot programmes, ensuring there are procedures for informed consent for participants; (iv) ensuring appropriate ethical review if pilot studies are defined as research, or compliance with health screening guidelines if they are defined as health screening; (v) in transfer from research studies to widespread roll-out across a whole community, ensuring fidelity to original protocol defined in research studies; and (vi) above all these issues, estimation of the balance between cost and effectiveness.

There are other technical issues in designing studies to evaluate screening and intervention to reduce conduct disorder in children and its possible aftermath. They include the following: (i) It is necessary to report results as numbers (or proportions) of participants meeting some criterion of improvement as a result of the intervention (rather than as aggregate data, often the mean scores on relevant rating scales, averaged over many participants). (ii) If randomization between intervention and comparison conditions is arranged, is this random assignment of individuals, or of groups (classrooms or schools)? (iii) Which type of intervention is most effective (for instance ones aimed to help children, parents or both; intervention delivered by home visits, in schools or at other venues). (iv) Beyond demonstration of the efficacy of an intervention (that is showing a major difference in controlled trials between groups with and without the intervention), it is important to examine intervening variables in order to understand the processes by which this is achieved.

Many studies have investigated this sort of intervention, and few will be cited here. However the above topics will be discussed generically, overall conclusions drawn about the current “state of the art”, and a few exemplary studies will be described, to show how such work is best done, combining scientific rigour with ethical probity.

(ii) *Problems in design.* After early interest in this type of intervention, various studies showed efficacy in research conditions. However, between initial definition of a population eligible for intervention, and final selection of participants recruited into a study, there is commonly substantial attrition, and potential for biased

sampling. For instance, in a study on long-term effects of home visits by nurses before a child's second birthday<sup>143</sup> on antisocial and criminal behaviour when the children reached adolescence, the following recruitment strategy was adopted: Initially 500 pregnant women from low socio-economic groups were approached, of whom 100 declined to participate. Other participants were invited from higher social strata, and the final cohort included 340 from low SES groups. Those who refused to join the project may have had more-than-typically severe problems compared with the cohort as a whole. In another study<sup>144</sup> of approximately 709 children eligible for the study (638 approached from nine centres, supplemented by the number expected from the tenth centre that refused to join), the number finally recruited into the experimental-plus-control groups was only 396 (=56% of those initially eligible). In these studies, persons eventually recruited are probably biased towards the less severe range of problems. Further attrition often comes before data on final outcomes are obtained, with the possibility of further biases. For instance in another study<sup>145</sup>, out of a large population approached in the general community, 364 met initial screening criteria (which represented about 401, after adjusting for those refusing initial screening). Of these, 332 gave consent, 306 were still available when contacted by research staff, 285 agreed to assignment to intervention or control groups, 259 provided data at the first evaluation, and, at the final evaluation, only 245 remained, about 61% of those initially eligible.

Bias in sampling compared to the general population may be introduced by the exclusion criteria used. Amongst exclusion criteria in one study<sup>146</sup> were that children already met criteria for a DSM diagnosis of Conduct Disorder or Pervasive Developmental Disorders, severe to profound mental retardation, and, amongst the children's parents, substance abuse, psychotic disorders, or pre-existing medical conditions. These exclusion criteria probably excluded many of the most severe or complex problems.

Many studies report outcome data only as mean values on rating scales for experimental and comparison groups, not as numbers, or proportion of individuals within each group meeting certain outcome criteria. These two ways of presenting results can be roughly matched, if one knows at what point on a rating scale, various percentile cut-offs lie for the relevant population. Occasionally this is done in studies of intervention outcomes for prevention of conduct disorders. For instance, in one study<sup>147</sup>, the mean initial score of on a Conduct Disorder scale was above the

---

143 Olds,D, Henderson,CR, Cole,R, Eckenrode,J, Kitzman,H, Luckey,D, Pettit,L, Sidor,K, Morris,P, Powers,J. (1998) Long-term effects of nurse home visitation on children's criminal and antisocial behavior. *JAMA*, 280, 1238-1244.

144 Webster-Stratton, C. (1998) Preventing conduct problems in head start children; strengthening parenting competencies. *Journal of Consulting & Clinical Psychology*, 66, 715-730.

145 Barrera,M, Biglan,A, Taylor,TK, Gunn,BK, Smolkowski,K, Black, C, Ary,DV, Fowler,RC (2002) Early elementary school intervention to reduce conduct problems: a randomized trial with Hispanic and Non-Hispanic children. *Prevention Science*, 3, 83-94.

146 Brotman,LM, Klein,RG, Kamboukos,D, Brown,EJ, Coard,SI, Sosinsky, LS(2003) Preventive intervention for urban, low-income preschoolers at familial risk for conduct problems: a randomised pilot study. *Journal of Clinical Child & Adolescent Psychiatry*, 32, 246-257.

147 Scott,S, Spender,Q, Doolan,M, Jacobs,B, Aspland,H, (2001) Multicentre controlled trial of parenting groups for childhood antisocial behaviour in clinical practice. *BMJ*, 323, 194-198.

97<sup>th</sup> percentile for the age group. After the intervention it fell to 82<sup>nd</sup> percentile, well within the normal range in the experimental group, but did not change in the comparison group, initially equally at risk. Nevertheless, 21% of children in the intervention group failed to improve by more than 0.3SD, and a third of this group still met criteria for Oppositional Defiant Disorder.

One rather surprising shortcoming is that many of studies apparently had no prior scrutiny by an ethics committee. Admittedly, there is generally some sort of consent process for families involved. In one study<sup>148</sup>, parents were paid a small honorarium for participation (e.g. US\$15 for two hours involvement). Lack of ethical scrutiny is an important omission, if large numbers of children and their families are randomly assigned to intervention or comparison conditions. It may be the consequence of this type of work falling under the heading of “social or policy research” rather than “health research”, even though, by using random assignment to different conditions, it is using a design typical of medical trials. Some studies did have proper ethical scrutiny. These include an impressive long-term study, sponsored by the World Health Organization starting in 1969<sup>149</sup>, part of a child mental health research program on the island of Mauritius, which adhered to Principles in the Declaration of Helsinki. Results showed that children who joined in an enrichment program at ages 3-5 years had lower scores for schizotypal personality and antisocial behaviour at age 17 years, and for criminal behaviour at age 23 years, compared with the comparison children. Beneficial effects of the intervention were greater for children showing signs of malnutrition at age 3 years, particularly for schizotypy at ages 17 and 23 and for antisocial behaviour at age 17.

(iii) *Effectiveness*. The central questions here are whether interventions are effective, what sort of intervention is most effective, targeted at which group, and under what conditions. There seems to be little doubt about the overall efficacy of intervention under research conditions, shown recently in a systematic review by the Cochrane collaboration<sup>150</sup>. The authors summary conclusions read as follows: “Behavioural and cognitive-behavioural group-based parenting interventions are effective and cost-effective for improving child conduct problems, parental mental health and parenting skills in the short term. The cost of programme delivery was modest when compared with the long-term health, social, educational and legal costs associated with childhood conduct problems. Further research is needed on the long-term assessment of outcomes.” Cost-effectiveness in mass screening and intervention programs, especially if intended to be nation-wide, such as *B4 School*, is discussed further below.

The question of which elements of intervention contribute to the efficacy, can be subdivided into several subsidiary issues. Two studies have compared intervention

---

148 Conduct Problems Prevention Research Group. (1999) Impact of the Fast Track Prevention Trial for Conduct Problems: I. The High-Risk Sample. *Journal of Consulting & Clinical Psychology*, 67, 631-647.

149 Raine, A, Mellinger, K, Kiu, J, Venables, P, Mednick, SA (2003) Effects of environmental enrichment at ages 3-5 years on schizotypal personality and antisocial behavior at ages 17 and 23. *American Journal of Psychiatry*, 160, 1627-1635.

150 Furlong M, McGiloway S, Bywater T, Hutchings J, Smith SM, Donnelly M. (2012) Behavioural and cognitive-behavioural group-based parenting programmes for early-onset conduct problems in children aged 3 to 12 years. *Cochrane Database of Systematic Reviews*, CD008225.

targeted at parents, children or both together. One of these<sup>151</sup> showed that, for some outcomes, (conflict management, other problem solving skills, interaction with friends), intervention was more effective if aimed at children than at parents. For measures of behaviour at home, intervention targeted at parents was usually more effective. In general however, a combination of the two was more effective than either by itself. The second study<sup>152</sup> had a massive drop-out before follow-up, making it unclear what conclusions were to be drawn.

A second issue is whether intervention based around schools is to be preferred to home visits, for instance by community nursing staff. Most studies upon which the Cochrane Collaboration review (see above) is based, were centred around schools. The issue of whether home visits are effective (or more effective than school-based interventions) is part of a wider debate about the effectiveness of home visits versus out-patient consultations in delivering health care generally, not only in psychiatry, but in other areas, such as paediatrics or old-age medicine. There seems little doubt about the value of home visits, so that medical staff can learn about patients' life style, establish rapport with a family (especially initially), and foster learning and training for doctors or trainees in real-life situations, especially for those, such as immigrant doctors, unfamiliar with the society and culture in which they practice. It may be most valuable for house-bound patients, such as the elderly and families with young children. People needing mental health care may fall into such groups. Usually such visits are highly appreciated by families who are visited. However, home visits for health care have fallen out of fashion in the last fifty years, and in psychiatry were scarcely considered until closure of asylums came on the agenda.

A long-term study already mentioned<sup>153</sup> in Elmira county, New York State, provided impressive evidence of effectiveness of home visits by nurses: They reduced child abuse, neglect, and associated outcomes in teenage years, such as delinquency, involvement with the justice system, drug and alcohol problems, or risky sexual behaviour. This is the only study to have shown this major beneficial effect. A review of the topic<sup>154</sup> highlights the differences between this successful program and others on the merits of home visits. The study involved nursing staff, not paraprofessionals (lay home visitors), and home visitors were involved with the families concerned both prenatally, and for two years after a child's birth. Although use of nursing staff was more costly per visit than use of paraprofessionals, costs savings eventually outweighed this by four times that of the intervention. However, as noted already, in this study there was substantial attrition, which may have led to sampling bias in favour of less severe problems.

---

151 Webster-Stratton,C., Hammond,M. (1997) Treating children with early-onset conduct problems: a comparison of child and parent training interventions. *Journal of Consulting Clinical Psychology*, 65, 93-109.

152 Larsson,B, Fossum,S, Clifford,G, Drgli,MB, Handegard,BH, Morch,W-T. (2008) Treatment of oppositional defiant and conduct problems in young Norwegian children. Results of a randomised controlled trial. *European Child & Adolescent Psychiatry*, 18, 42-52.

153 Olds,D, Henderson,CR, Cole,R, Eckenrode,J, Kitzman,H, Luckey,D, Pettit,L, Sidor,K, Morris,P, Powers,J. (1998) Long-term effects of nurse home visitation on children's criminal and antisocial behavior. *JAMA*, 280, 1238-1244.

154 MacMillan, HL (2000) Child maltreatment: What we know in the year 2000. *Canadian Journal of Psychiatry*, 45, 702-709.

Several studies try to identify in more detail the processes by which benefits of intervention are achieved. All effects are complex ones involving improved social interactions. One effect was produced immediately through a child's behaviour, the peer groups with whom s/he chooses to interact then favouring further development of more positive strategies of behaviour<sup>155</sup>. A second process, identified in long-term study of home visits<sup>156</sup>, was that, by effects on a child's behaviour, episodes of maltreatment in the family were less likely. In a study where children, parents and teachers were all involved in the intervention<sup>157</sup>, not only did children's behaviour contribute to overall improvement, but so did the parents' parenting skills, and the teachers' classroom skills. A fourth study<sup>158</sup>, involving a six-year follow-up, showed that reduction in symptoms of Oppositional Defiant Disorder was mediated partly by improved social skills in children and adolescents, but also by their learning a more effective personal discipline.

With regard to the whole field of screening and intervention to prevent conduct disorder and associated problems, a recent review<sup>159</sup> takes a hard look at whether such programmes can be effective if rolled out on a mass scale. It draws attention to the difficulty in maintaining fidelity to original protocols proven to be effective in a research study when implemented on a larger scale. Sometimes these imperfections in fidelity are unintended consequence of the increased scale of the intervention itself. For instance, a study in California became so large that it involved recruiting many new teachers, which could not be done without lowering their standards of training and later performance. The review also points out that, while screening instruments may be relatively effective in detecting problem children in groups at highest risk, these groups are a minority of the population to be screened. When the size of the respective risk-groups is multiplied by their respective relative risk, it is clear that most of the problems do not arise from groups at highest risk. For instance in the study cited in their review, part of the Conduct Problems Prevention Research Group, started in 1992<sup>160</sup>, the group at highest risk (3% of the population) had a prevalence of conduct disorder of 41%, while the remaining 97% had a prevalence of only 6%. However the highest risk group, identified by the screening method, made up only 17.5% of all cases of conduct disorder (product of cohort size and

---

155 Vitaro,F, Brendgen,M, Pagini, L, Tremblay,RE, McDuff,P. (1999) Disruptive behavior, peer association, and conduct disorder: Testing the developmental links through early intervention. *Development & Psychopathology*, 11, 287-304.

156 Eckenridge,J, Zielinski,D, Smith,E, Marcynyszyn,LA, Henderson,CR, Kitzman,H, Cole,R, Powers,J, Olds,DL (2001) Child maltreatment and the early onset of problem behaviors: Can a program of nurse home visitation break the link? *Development & Psychopathology*, 13, 873-890.

157 Webster-Stratton,C, Reid,MJ, Hammond,M. (2001) Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start. *Journal of Clinical Child Psychology*, 30, 283-302.

158 Bernat,DH, August,GJ, Hekter,JM, Bloomquist, ML. (2007) The early rises preventive intervention: Testing for six-year outcomes and mediational processes. *Journal of Abnormal Child Psychology*, 35, 605-617.

159 Dodge, KA (2009) Community intervention and public policy in the prevention of antisocial behavior. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 50, 194-200.

160 Flanagan,KS, Bierman,KL, Kam,CM (2003) Identifying at-risk children at school entry: The usefulness of multibehavioral problem profiles. *Journal of Clinical Child & Adolescent Psychiatry*, 32, 396-407.



relative risk:  $0.03 \times 0.41 = 1.23\%$  vs  $0.97 \times 0.06 = 5.82\%$ ;  $1.23/5.82 = 0.175$ ). The authors comment as follows, about the success of the program involving home visits by nurses: “Although one can be in awe of the scope of this implementation, the numbers reveal that no community is likely to receive full penetration to all qualifying mothers, and no report has yet been issued regarding the effect on population prevalence of child abuse or conduct disorder in any community.” They conclude as follows: “Given the modest evidence and the challenges in community implementation, it is not clear that population-level impact will occur even for this best-evidenced and best-funded program.”

(iv) *Three exemplary studies.* In the rest of this section, a description is given of three interesting projects, all screening for risk of, and intervening to prevent developing conduct disorder and related problems. None is perfect, but all have strong features along with corresponding weaknesses. The first<sup>161</sup>, conducted in inner-city Montreal used the Social Behavior Questionnaire. There were 1161 children at risk, mean age 6.1 years, identified by kindergarten teachers, as scoring above the 70<sup>th</sup> percentile on a disruptiveness scale. After applying exclusion criteria and with attrition, this number fell to 879. The intervention, which extended over 2 years, included social skills training for these children, aimed at promoting change in behaviour towards peers, yielding more social acceptance, and less inclination towards antisocial peers. Training was arranged in groups of 4-7 children, with a ratio three pro-social to one disruptive child in each group. Parent training in effective child-rearing was also arranged. Three groups were compared, high-risk children with and without intervention, and low-risk children. Outcomes at age 24y documented as rates of high-school graduation and criminal conviction, are given in Table VI. After controlling for parental occupation and initial level of disruptiveness, differences between groups for high-school graduation reached a significance level of  $p < 0.05$ , and, for criminal record,  $p = 0.06$ . The Intervention group and the Low-Risk group were not statistically different on either measure, but there were highly significant ( $p < 0.0001$ ) differences on both measures between the Control and the Low-Risk groups.

**TABLE VI: Outcomes in Montreal Study as a Function of Intervention.**

	Control	Intervention	Normative
High sch grad:	56/174 (32.2%)	31/68 (45.6%)	340/637 (53.4%)
Criminal record:	59/181 (32.6%)	15/69 (21.7%)	104/646 (16.1%)

Unlike use of the SDQ in *B4 School*, this *was* declared as research, although there is no mention of either ethical scrutiny or informed consent by parents. Unlike *B4 School*, the intervention, follow-up data collection, and analyses were well thought out, as was control for possible covariates. In addition, the scales used had a single purpose, not a spurious claim to identify several different problems. Scaling

---

161 Boisjoli,R, Vitaro,F, Lacourse,E. (2007) Impact and clinical significance of a preventive intervention for disruptive boys. *British Journal of Psychiatry*, 191, 415-419.

results up to the total at-risk cohort of 242 (i.e. 68+174), adverse outcomes for graduation would have occurred without intervention in 78, and with intervention in 110. For criminal record the numbers (out of a total of 250 [69+181] without intervention) would have been 81.5, and with intervention would be reduced to 54.25. Given that there had already been some selection (~20%; including, as an exclusion criterion that parents had less than 14 years of schooling) which may have excluded highest-risk children, where effects of intervention may have been less, the effects may have been even less overall. In summary, significant effects were seen long-term for the intervention, but not ones which were clearly applicable in a mass screening program.

The second study<sup>162</sup>, from Bangor in North Wales, is currently in progress. The design is described, but results are not yet available. About two hundred and forty children (aged 4-8 years) from Bangor rated by their teacher as above the “borderline cut-off” on the SDQ, as well as their parents, will be recruited. Randomisation is to be by individuals within schools, in a 1:1 ratio between intervention and waiting list control. (The control children *are* to receive the intervention after the first follow up.) Phase 1 comprises three data collection points - baseline and two subsequent points eight months apart. Phase 2 includes baseline and first follow-up. The Programme includes: learning school rules; understanding, identifying, and articulating feelings; problem solving’ anger management; how to be friendly; how to do your best in school. Outcomes are: Change in child social, emotional and behavioural difficulties; teacher and parent mental wellbeing; child academic attainment; child and teacher school attendance. Intervention delivery will be assessed for fidelity.

Although not yet implemented, there are many strengths in this design: (i) It uses SDQ total, with “borderline” score as cut-off, and no attempt to use subscale scores. (ii) It is small scale, (iii) It is declared as a research study, with the rationale carefully defined, as well as secondary objectives. (iv) It specifies what parents understand and give consent to; (v) Even when it is acknowledged to parents that a child has a problem, the objective is open and supportive to parents; (vi) It was subjected to ethical scrutiny. (vii) It was registered in the UK Clinical Trials Network. However, as described only short-term outcomes are covered.

The third study<sup>163</sup> was carried out on Lambeth (central London). Nine hundred and thirty-six 6-year-olds were screened for antisocial behaviour, of whom 279 (41%) were above the cut-off for antisocial behaviour (typical of this highly deprived area). Of these 128 were eligible and gave consent, eventually reducing to 112 high scorers, who were randomised to parenting groups (held in schools) or a control group; 109 were followed up a year later. In terms of SDQ scores these 112 were similar to 1567 non-participants. The commonest reasons for non-participation were “too busy” due to jobs or courses. Intervention lasted 28 weeks, with components addressing both child behaviour and literacy. Fidelity to the protocol was emphasised by training of

---

162 Bywater, T, Hutchings, J., Whitaker, C, Evans, C, Parry, L. (2011) The incredible years therapeutic dinosaur programme to build social and emotional competence in Welsh primary schools: study protocol for a randomised controlled trial. *Trials (electronic resource)*, 12, 39.

163 Scott, S, Sylva, K, Doolan, M, Price, J, Jacobs, B, Crook, C, Landau, S. (2010) Randomised controlled trial of parent groups for child antisocial behaviour targeting multiple risk factors: the SPOKES projects. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 51, 48-57.

therapists and weekly supervision. Controls received an information helpline. Assessment of conduct problems was by parent interview, parenting by direct observation and child reading by psychometric testing.

Compared to control children whose behaviour did not change, conduct problems in the intervention group reduced by 0.52 SD ( $p < 0.001$ ), dropping from the 80th to the 61st percentile; oppositional-defiant disorder (ODD) halved from 60% to 31% ( $p = 0.003$ ). ADHD symptoms reduced by 0.44 SD ( $p = 0.002$ ), and reading age improved by six months (0.36 SD,  $p = .027$ ). The programme cost £2,380 (\$3,800) per child. Written consent was obtained from parents, and the local research ethics committee approved the project. This study shows that effective early intervention to improve functioning of children with antisocial behaviour is practically feasible by targeting multiple risk factors and emphasising fidelity to the protocol. However, in view of high rate of attrition it is unclear whether or how the programme could be delivered to all who would benefit from it, and, since follow-up was only over 1 year, it is not clear what the long-term impact might be.

The overall conclusion from this review is to acknowledge that very effective intervention *is* possible in principle. However, as often found in health screening programmes, those most in need are the ones hardest to reach. Reiterating the comment quoted above from a 2009 paper: “It is not clear that population-level impact will occur even for this best-evidenced and best-funded program.” It may be that the best way to use the research in this area is not in deliberate screening and intervention within paradigms of preventive medicine, but to use the knowledge, and perhaps some of the screening instruments used, in routine clinical services. As mentioned in section B, effective programmes need to achieve a balance between mass screening/intervention programmes, ones targeted at specific groups, and ones fully integrated within clinical services.

(iv) *Implications for Policy: Confusion of Health, Justice and Social Development Agencies.* In the background to use of the SDQ to screen for, intervene and prevent the sequelae of conduct disorder in children is an older debate, supposedly using scientific methods to predict criminality. The idea that criminality, being dependent on a person’s intrinsic makeup, genetics or biology, can be understood scientifically, has a long, and not-always glorious history, with Cesare Lombroso (1835-1909), an Italian criminologist of the late nineteenth century as a dominating figure. In the last generation this has morphed into quasi-medical approaches to crime as “risk assessment and intervention” rather than “innocent until proved guilty”. So, when Tony Blair was running for office he used the slogan “Tough on crime; tough on the causes of crime”. How did it pan out, when New Labour was in office? One example was the attempt to get adopted as an official psychiatric diagnosis “Dangerous Severe Personality Disorder”, this to be used to detain people thought to be at risk, but who had committed no crime. It had neither a psychiatric nor a legal foundation. No doubt it pleased part of his electorate, but fortunately, was stopped by knowledgeable community voices combining with the psychiatric profession. The latter insisted that they were being asked to do something, by way of prediction, which was, in practice, impossible. Hidden under all the detail, there is something similar in the use of the SDQ as a component of *B4 School*.

Since WWII and the first randomised controlled trial (RCT) of streptomycin, such trials have become commonplace in medicine, initially controversial, but now usually

safe, with appropriate ethical safeguards. They are sometimes held up as the gold-standard for proof of the efficacy or otherwise of an experimental treatment, although it is best if such evidence is supplemented with that from a variety of other sources. However, for studies on intervention to reduce conduct disorder and its sequelae, we are in the area of social or policy research not health research, although still using the RCT design developed for medical research. The place of ethics, and the role of ethics committees for such “experimental sociology” is not well established as it is in medicine. One may ask if experimental sociology is ever permissible from an ethical point of view, and if so, under what conditions. For predicting antisocial or criminal tendencies, to which “innocent until proven guilty” was hitherto a core assumption in the justice system, we seem to be using concepts from medical research, properly used there to protect people from serious health problems, but used here, indirectly and covertly to protect society from criminality. Is this health? Is it justice? Is it democracy?

## **G: How have the SDQ, and the Data it Gives Been Used?**

[I] *Hawke’s Bay Report, 2010*. The Ministry of Health launched the *B4 School Check* in February 2008. After the system had been in operation for ten months, major stakeholders in the Hawke’s Bay Primary Health Organization collected data on its use in that region, as reported in November 2010<sup>164</sup>. In these 10 months, 1848 checks were carried out (84% of the eligible cohort of 2180)<sup>165</sup>. The percentage of the cohort checked in each socio-economic quintile declined from quintile 1 to quintile 5, but was still as high as 75% in quintile 5 (for which 36% of checks were carried out). An average of 5.5% of children checked were referred for further checking/help, the proportion increasing across SES quintiles, reaching ~7% in quintile 5. The overall percentages referred for each part of the *B4 School Check* were as follows: Dental: 13.5%; Hearing: 11.5%; Vision: 11.1%; Immunisation: 7.2%; PED: 7.0%; SDQ: 3%; General Health: 1.5%; Growth: 0.4%<sup>166</sup>. In quintiles 1 and 2, about 1% of children were referred from SDQ scores, in quintile 3: ~2%, and for quintiles 4 and 5: approximately 4% (and 8.5 based on PEDS results). As a percentage of all referrals, 5% were based on SDQ scores, and 13% on PEDS results. Feedback from nurses suggested that SDQ referrals under-represented the number with behaviour/emotional problems. Nevertheless, NGOs concerned with social work, counselling or parenting programmes reported accelerated demand that exceeded their resources, and the Child Development Unit in Hawke’s Bay also reported increased workload. The authors comment: “The apparently low referral rate for behaviour of 3% is approximately half of that expected, but could be explained by some referrals for behaviour from PEDS, which is administered first in Hawke’s Bay.” However, the relevant three

---

164 Wills,R, Matthews,KM, Hedley,C, Freer,T, Morris,H, (2011) Improving school readiness with the Before School Check: Early Experience in Hawke’s Bay. *New Zealand Medical Journal*, 123 (issue no 1326), 47-58.

165 No separate figures were given for the different sections of *B4 School*, implying that the consent form covered the whole check, with no differentiation between its parts.

166 These, and other values are approximate, having been read from the figures in this paper.

items in the PEDS<sup>167</sup> are hardly as specific as those in the SDQ. Moreover, PEDS referrals did not differ much across SES quintiles, unlike the SDQ referrals. If the PEDS was picking up problems which otherwise would have been picked up by SDQ, one might have seen a similar trend. It should be noted also that in the *Practitioner's Handbook* (p.37), the role of the PEDS is stated to be mainly for surveillance, not screening, and so should not have been used as implied in this report. The authors also comment "Narrative feedback from nurses suggests some resistance to the SDQ from parents and nurses, however it appears to be gaining acceptance as nurses become more confident in its use." "There . . . remains scepticism about the SDQ in particular and engagement remains patchy". Nevertheless, the use of the SDQ drew attention to some problems of hard-to-reach families, who could be helped via home visits.

**[II] Editorial in NZMJ, 2010.** Based on the data from Hawke's Bay, an Editorial appeared in the same number of New Zealand Medical Journal<sup>168</sup>, where we read: "The *B4 School Check* introduced validated, evidence-based behavioural and developmental questionnaires". "As a result of the significant numbers of *B4 School* checks now being delivered through PHO-based primary care, general practitioners and Well Child/Tamariki Ora nurses are beginning to work more closely, and the concept of broader team-based care is now beginning to emerge." In Hawke's Bay: "The DHB's decision to contract a PHO-based provider to deliver the *B4 School* check gave an opportunity for general practice to make strong professional relationships with a range of community-based Well Child providers. It also enabled general practitioners to develop their role in the provision of a universal Well Child/Tamariki Ora programme." "The other key decision made by the DHB was the implementation of a Clinical Advisory group, chaired by the Director of Paediatrics. This group provided clinical governance, and enhanced inter-sectoral collaboration by bringing together the major child health, education and well-being groups from across the Bay. The collaborative nature of the Clinical Advisory group led to high levels of parental acceptance of the programme." "If the *B4 School Check* and the new Well Child programme are implemented nationally with as much commitment as in Hawke's Bay, then the lifetime opportunities for all New Zealand children can only be improved."

**[III] Two Recent Sources of Data.** These were preliminary results from a single region of New Zealand. How has the *B4 School Check* fared since then, and in particular how has the SDQ instrument been taken up, and its scores utilised? Two other, more recent sources of information are available: One is an e-mail exchange between the author and Leonie McCormack, of the Ministry of Health (7.02.2013), giving some statistics about national-wide use of the SDQ up to the time of writing. The other is the newsletter *B4SC Quality letter, August 2012*, giving figures for the

---

167 The three relevant questions (out of ten) in the PEDS are as follows: 1: Please list any concerns about your child's learning, development and behaviour; 6: Do you have concerns about how your child behaves? 7. Do you have any concerns about how your child gets along with others?

168 Tuohy,P. (2010) The B4 School Check – addressing the new morbidity in child health. *New Zealand Medical Journal*, 123 (issue number 1326), 19-21.

uptake, and referral rates, of the whole *B4 School Check* for the most recent quarter, broken down by section of the check, and District Health Board.

From the MoH correspondence, the following data were obtained: At the time of writing, the numbers screened with the parent version of SDQ was 180,198, and with the teacher version, 92,339. Of those approached for completion of the SDQ, there were 3,252 refusals from parents, and 46,566 from teachers. Adding those screened to those refusing, the totals approached amounted to 183,450 for the parent version, of whom 1.8% refused, and 139,505 for the teacher version, of whom 33.5% refused. Thus, teachers are far more reluctant to use the SDQ than parents. Of those screened with the SDQ, 3291 (1.8%) with the parent version, and 1,760 (1.9%) of with the teacher version were referred for further assessment/treatment.

While the SDQ was intended for screening rather than diagnosis, it has within its structure the ability to screen for several different sorts of difficulty, and is presented as though such differentiation was one of its objectives. However, no data appear to be available on numbers or proportions of children rated ‘borderline’ or ‘abnormal’ on each of its main subscales.

While health screening should be linked to deliverable interventions or outcomes, there appears to have been no attempt to collect data on referral pathways, or other outcome measures for those rated “borderline” or “abnormal” via the SDQ. No doubt this is explained by difficulty of follow-up. Gathering outcome data is seen as an important part of modern health administration, yet here, “outcome” seems to be no more than a statistic of numbers/proportion of children screened and referred.

From the *B4SC Quality Newsletter*, it is clear that there are massive variations between DHBs in the percentage of children screened with the SDQ, parent version, who are rated “borderline” (1.6-7.9%; mean across all DHBs 5.0%), or “abnormal” (1.6-7.9%; mean 4.2%). For the teacher version, the corresponding values are: “borderline”: 0.7-6.5% (mean 3.7%), and “abnormal”: 0.0-6.5% (mean 2.6%). Parents tended to rate more severely on the SDQ than teachers, but more consistently across regions, and with much more consistent ratio of borderline to abnormal. Apart from the vast variation across regions, of those screened, a much smaller percentage (on average) are found “abnormal” than in studies overseas, and of this small fraction, less than one half are referred. Of those who are screened and found “abnormal” the variation between DHBs on the percentage referred is even larger (for the parent version: 0-94%; mean 21%; for the teacher version: 0-94%; mean 28%). There also appears to be inconsistency on whether informed consent is given by parents to *B4 School* as a “package deal”, or alternatively that consent can be refused for some parts while accepting other parts. This may depend on details of the way *B4 School* is administered in each DHB.

Why the high refusal rate, especially amongst teachers? There may be suspicion of the test itself by teachers. There may also be some disregard of this part of the test by the health practitioners who actually supervise delivery of this test. One way to detect this would be to compare the total numbers screened via the SDQ and in other parts of *B4 School* (probably a larger number, but how much larger is uncertain). In the *B4SC Quality Newsletter, August 2012*, total numbers screened are given for most sections of *B4 School*, but not for the SDQ (except for the teacher version). There may also be neglect of this test, or referral based on it if scores are abnormal, and of reporting of the outcomes, because it is known that there is either no referral pathway,

or not one that could cope with resulting demand. On an individual level however, there is anecdotal evidence of people who were significantly helped by this part of *B4 School*. It is probably true that the most energetic implementation of the program including the SDQ (as in Hawke's Bay) may be paying off. Whether it can ever be effective across the whole country, is uncertain.

**[IV] Policy Implications:**

(i) *Does Use of the SDQ Conform to Health Screening Guidelines?* These guidelines were listed in section B. On the *first* criterion, that the condition be a suitable candidate for screening, one can give no clear answer. This the underlying questions raised by this report. On the *second*, that there is a suitable test for the condition, the answer is a cautious "yes", but the best method is not the SDQ, and may be more labour- or resource-intensive than the SDQ. On the *third* and *fourth* criteria, that there be an effective and accessible treatment or intervention, supported by high-quality evidence, especially randomised controlled trials, the answer is a cautious "yes" in principle, but major uncertainties prevail on whether it could ever be a mass screening and intervention programme in practice. On the *fifth* criterion, that potential benefit outweighs potential physical and psychological harm, great caution is urged, especially if (as seems to be the case here), the SDQ was introduced in a somewhat surreptitious manner, its motives not made entirely clear to recipients. On the *sixth* criterion, that the health care system can support all necessary elements of the screening pathways, including diagnosis, follow-up and programme evaluation, the answer is a decisive No. The *seventh* criterion, that social and ethical issues are considered, is answered with a decisive No. Lastly, cost-benefit considerations of use of the SDQ are shrouded in mystery.

(ii) *Governance issues: Danger of Multi-Agency, Multi-Government Coordination.* The *B4 School* was planned by one government, and implemented by another of supposedly-opposite political hue. For behavioural and emotional screening, it involved ministries of health, justice, education and social development, as well as a variety of academic researchers, and the initiative obviously involved multi-agency coordination. Of course, coordination is needed between government departments, particularly so that economic and social policies can be reconciled. There seems to be a long way to go here, and much good to come from such coordination. Nevertheless, there is also need for some checks on cross-department coordination. In the context of issues discussed here, hard questions should be asked: Who is in control? How are decisions taken? Who is pulling the strings? What is the chain of accountability within government, and to the wider electorate? In addition, for an initiative dealing with highly personal, and potentially sensitive information, one should ask: What limits are there on data-sharing between databases collected by the several ministries? Are there conflicts of interest? This issue was highlighted by a recent announcement (8<sup>th</sup> April) of a soon-to-be-delivered policy from the Ministry of Social Development, under which beneficiaries could have their benefits cut in half if they or their children had not fulfilled health checks and other requirements<sup>169</sup>. On 14<sup>th</sup> April there was a further announcement of success in detecting benefit fraud by matching data collected by Inland Revenue and MSD. On 11<sup>th</sup> April 2013, we heard a government

---

169 Radio New Zealand National Program News, 8.04.2013.

announcement that it was moving ahead with a massive program of inter-agency data sharing. Claims were made that all personal data would be securely anonymised, but statements made three days earlier and three days later from MSD suggest that data matching is not necessarily anonymous<sup>170</sup>, to say nothing of well-known and frequent leaks of confidential information from government websites. The interviewer asked: “Who is accountable?” and the apparent answer was “Head of the State Services Commission”, an unelected position. Quite apart from lack of accountability to the electorate, the growth of such multi-agency consortiums might be seen as another way in which transparency of collective decision-making by cabinet, in our style of government is undermined. Across history, other governments have adopted a rather different style. There is a word for their style: Totalitarianism. It is therefore perhaps fitting to finish by quoting Thomas Jefferson’s famous line: “*The Price of Liberty is Eternal Vigilance*”.

## **H: Implications for Australia?**

The behavioural and emotional development part of *B4 School*, has implications for Australia as well as New Zealand. The present author was alerted to this by a visit from Professor Allen Frances, in July, 2012. His over-riding concern was diagnostic inflation, but one of his sharper focuses was over-diagnosis of mental disorders in children. His visit to New Zealand followed a longer visit to several centres in Australia. At the workshop on Psychiatric Diagnosis in Wellington<sup>171</sup>, it was clear that Australian health services intend to implement something similar across the whole country, and were watching with interest the use of the SDQ in *B4 School*. Concern about this has already been voiced in Australia<sup>172</sup>. A recent review in *Medical Journal of Australia*<sup>173</sup> concludes: “Despite good intentions, further research and planning are needed before starting large-scale, population social-emotional screening of 3-year-old children in Australia.”

## **I: Recommendations.**

---

170 It is interesting how almost identical issues emerge in the news, almost simultaneously in New Zealand and Britain. In the context of the above news items, the reader is referred to the following two news items in *Guardian*: 29.04.2013: Rajeev Syal: *How HMRC treated its Goldman Sachs tax deal whistleblower as a criminal*. 30.04.2013; Shiv Malik: *Jobseekers made to carry out bogus psychometric tests*.

171 R.Miller, Report and Commentary on Workshop on ‘Diagnosis in Psychiatry’, Brentwood Hotel, Wellington, 16th July, 2012. See: [www.robertmiller-octspan.co.nz](http://www.robertmiller-octspan.co.nz) (OCTSPAN side).

172 Jureidini,J, Raven,M. (2012) Healthy Kids Check: Lack of Transparency and misplaced faith in the benefits of screening. *Australian & New Zealand Journal of Psychiatry*, 46, 924-927; Kowalenko,NM (2012) The three year old social and emotional well-being check is good news. *Australian & New Zealand Journal of Psychiatry*, 46, 928-931.

173 Daubney,MF, Cameron,CM, Scuffham,PA. (2013) Changes to the Healthy Kids Check: Will we get it right. *Medical Journal of Australia*, 198, 475-477.



- (i) All health screening programs should conform to the 2003 Ministry of Health guidelines.
- (ii) The SDQ should be rigorously evaluated with respect to the 2003 guidelines before any further use is made of it.
- (iii) Government officials who design such programs need better understanding of scientific methodology, its strengths and limitations, and the ethical and practical issues involved.
- (iv) There should be greater awareness by administrators developing health screening programs of the difference between screening which is scientifically valid at a population level, and that which is scientifically and ethically valid, and valid in terms of cost-benefit analysis, when applied to individuals.
- (v) Early intervention and programs for preventive medicine which involve mass screening need careful scrutiny, especially for disorders with multiple causes or involving collection of sensitive data, with cost-benefit considerations always in mind.
- (vi) It is a concern that data are being collected about large numbers of individuals that is potential sensitive and may be inaccurate, when often there is little chance of responding fully to increased demand for services to which such collection may lead.
- (vii) Even if the SDQ were to be properly validated for the age group, the extent of dissemination of data collected with this instrument requires careful consideration, public debate, and even legislation to prohibit breaches of confidence.
- (viii) Procedures for obtaining truly informed consent need to be more robust.
- (ix) There should be sufficient staff, adequately trained in child health and child mental health, to implement the B4School program, or whatever it becomes.
- (x) It is important to avoid pressures to merge health and justice agendas in policy on mental health, except within the highly specialised area of forensic psychiatry.