


Screening in early childhood for risk of later mental health problems: A longitudinal study

Jake M. Najman^{a, b, *}, Michelle A. Heron^a, Mohammad R. Hayatbakhsh^a, Kaeleen Dingle^a, Konrad Jamrozik^a, William Bor^c, Michael J. O'Callaghan^c and Gail M. Williams^a

^aSchool of Population Health, University of Queensland, Herston Road, Herston QLD 4006, Australia

^bSchool of Social Science, University of Queensland, St Lucia, QLD 4072, Australia

^cMater Misericordiae Children's Hospital, South Brisbane, QLD 4101, Australia

 Corresponding author. Address: School of Population Health, University of Queensland, Herston Road, Herston QLD 4006, Australia. Tel.: +61 7 336 55180; fax: +61 7 336 55509.

Abstract

Depression in childhood or adolescence is associated with increased rates of depression in adulthood. Does this justify efforts to detect (and treat) those with symptoms of depression in early childhood or adolescence? The aim of this study was to determine how well symptoms of anxiety/depression (A-D) in early childhood and adolescence predict adult mental health. The study sample is taken from a population-based prospective birth cohort study. Of the 8556 mothers initially approached to participate 8458 agreed, of whom 7223 mothers gave birth to a live singleton baby. Children were screened using modified Child Behaviour Checklist (CBCL) scales for internalizing and total problems (T-P) at age 5 and the CBCL and Youth Self Report (YSR) A-D subscale and T-P scale at age 14. At age 21, a sub-sample of 2563 young adults in this cohort were administered the CIDI-Auto. Results indicated that screening at age 5 would detect few later cases of significant mental ill-health. Using a cut-point of 20% for internalizing at child age 5 years the CBCL had sensitivities of only 25% and 18% for major depression and anxiety disorders at 21 years, respectively. At age 14, the YSR generally performed a little better than the CBCL as a screening instrument, but neither performed at a satisfactory level. Of the children who were categorised as having YSR A-D at 14 years 30% and 37% met DSM-IV criteria for major depression and anxiety disorders, respectively, at age 21. Our findings challenge an existing movement encouraging the detection and treatment of those with symptoms of mental illness in early childhood.

Keywords: Childhood; Adolescence; Adult; Psychopathology; Screening

1. Introduction

There is now substantial evidence that some mental health problems may have an early age of onset ([Costello et al., 2006] and [Dierker et al., 2001]) and high levels of recurrence over the life course ([Birmaher et al., 1996] and [Rutter et al., 2006]). Cross-sectional studies involving respondent recall are supported by longitudinal studies confirming the association between early age of onset and recurrence of mental health problems over the adult life course. Using maternal reports of psychopathology at age 13 and an interviewer-rated checklist at 24 years of age, Lynam et al. (2007) found a correlation of $r = 0.31$ between adolescent and young adult mental health scores. Repeated assessments of anxiety and depression from childhood (age 11 years) to age 32 years using data from the Dunedin birth cohort also suggest that both anxiety and depression are frequently characterised by adolescent onset and recurrence over the life course (Moffitt et al., 2007). Furthermore, there is some evidence indicating that early age of onset may predict a more chronic course for depression ([Kessler et al., 2001] and [Kovacs, 1997]). These findings raise the possibility

that young children and adolescents might be screened for symptoms of mental illness with the aim of recruiting those children with high scores into treatment and/or prevention programs.

The case for early intervention fundamentally depends on the extent to which low age of onset predicts significant ill-health in adulthood. Many early intervention and treatment programs require the screening of children to identify those at particular risk. Such approaches have intuitive appeal but, as Wilson and Jungner (1968) noted almost 40 years ago, there is a long and demanding list of criteria that should be met before screening programs are established. Three particular issues confront any advocates of screening children to identify and intervene with those at high risk of later problems with mental health. First, there is the need for a screening process or instrument that has acceptable validity, that is, both high sensitivity (the capacity to identify individuals who subsequently experience a mental illness) and high specificity (the capacity to exclude correctly those individuals who subsequently do not experience a mental illness). Specificity may be particularly important in such a stigmatized area as mental health, as mislabelling a healthy child as being at high risk could itself carry adverse consequences, even if any subsequent preemptive intervention was entirely harmless.

The second issue concerns the timing of the screening assessment or assessments, given evidence that there are sensitive periods in a child's development when exposure to adverse environments can initiate changes that have long term consequences for mental health. Logically, screening would follow rather than precede such periods. Third, having assessed children with an appropriate instrument at appropriate time(s), one must have available and be able to deliver an intervention that reduces later mental health problems among those deemed to be at high risk of such problems.

The present paper deals especially with the first two of these issues, the validity of a potential screening instrument and the timing of possible assessments. In addition, this study explores the positive predictive value (PPV); that is the proportion of the individuals with a positive screening test who later develop a significant problem with their mental health. Finally, this study examines the proportion of individuals who are labelled as being at high risk but who do not meet the criteria for mental health disorders by 21 years of age (false positives).

2. Materials and methods

2.1. Participants

We used data from the Mater-University Study of Pregnancy (MUSP), a prospective longitudinal study of a consecutive cohort of individuals born in Brisbane, Australia between 1981 and 1983 at a major public hospital (Mater Misericordiae Hospital). Recruitment procedures for the larger study have been described elsewhere ([Hayatbakhsh et al., 2007] and [Najman et al., 2005]). Of the 8556 mothers initially approached to participate in the study 8458 (98.9%) agreed. The cohort consists of 7223 women who delivered a live singleton baby who was not adopted out. This represents 87% of all women who attended the antenatal clinic during the study period. Mothers completed questionnaires at their first antenatal clinic visit, 3–5 days after the birth, 6 months after the birth, 5 years, 14 years and 21 years after the birth. Children completed their own questionnaires at 14 and 21 years. Due to a shortage of funding, only a subset of young adults ($n = 2563$) were administered the Composite International Diagnostic Interview, computerized version 2.1 (CIDI-Auto) at 21 years (World Health Organization, 1997). The average age of offspring participants was 20.4 years ($SD = 0.8$), 51.0% were female, 24.6% had tertiary education, 54.6% had completed high school and 20.8% had some primary or secondary school education.

2.2. Measures

Mothers completed a modified version of the Child Behaviour Checklist (CBCL) at the 5- and 14-year follow-ups, and children completed the Youth Self Report (YSR) at 14 years ([Achenbach, 1991a] and [Achenbach, 1991b]). These widely used and validated instruments served as screening tests for later mental health problems. They provide standardized checklists of child and adolescent behaviour problems and competencies. The CBCL is designed for completion by parents of children aged 4–18, and generates scores for eight syndromes: withdrawal, somatic complaints, anxiety/depression (together constituting the Internalizing Scale), delinquent behaviour, aggressive behaviour (together constituting the Externalizing Scale), social problems, thought problems and attention problems. A total problem score is derived by summing the individual item scores. In this study the Internalizing (Cronbach's $\alpha = 0.76$) and total problems (T-P) (Cronbach's $\alpha = 0.90$) scales were used at 5 years, and the anxiety/depression (A-D) (Cronbach's $\alpha = 0.85$) and T-P (Cronbach's $\alpha = 0.95$) scales were used at 14 years. The items completed at the 5-year follow-up, while used to measure internalizing symptoms, largely comprised symptoms of A-D.

The modified (short form) of the CBCL completed by mothers at 5 years included 33 of the 113 items from the original scale. Items selected assessed the most commonly occurring behaviours in 5-year-old. Furthermore, respondents in our study rated items as occurring 'often', 'sometimes' or 'never' rather than on a three-point scale ranging from 0 – 'not true' to 2 – 'very true' or 'often true', as described in the original scale. Factor analyses and reliability estimates of sub-scales produced results consistent with Achenbach's data ([Achenbach, 1991a] and [Najman et al., 2001]). In addition, a sample of 76 parents whose 6-year-old children were at school also completed the long form of the CBCL. There were very strong correlations (externalizing $r = 0.94$, internalizing $r = 0.89$) between the short and full forms of the CBCL (Najman et al., 1997).

The YSR is based on the CBCL and obtains self-reports from 11- to 18-year-old. These checklists have been shown to be reliable and arguably valid indicators of problem behaviour ([Achenbach, 1991a] and [Achenbach, 1991b]). The complete A-D scale (Cronbach's $\alpha = 0.84$) and T-P scale (Cronbach's $\alpha = 0.94$) were used and respondents were given the modified options of 'often, sometimes, and rarely/never'.

The CIDI-Auto is a structured interview that assesses mental disorders, both lifetime and current, according to the criteria of the ICD-10 (World Health Organization, 1993) and DSM-IV (American Psychiatric Association, 1994). The instrument contains 276 questions about potential symptoms and assesses symptom severity, help-seeking behaviour, psychosocial impairments and other episode-related information. The CIDI-Auto is a fully computerized version of the standard CIDI, which can be administered either by an interviewer or completed by the respondent. The data from the CIDI are entered into a scoring program which gives output according to the diagnostic criteria satisfied. The CIDI-Auto has good inter-rater and test-retest reliability (Peters et al., 1998) and has acceptable validity (Peters and Andrews, 1995). We used the CIDI-Auto to assess lifetime diagnoses of DSM-IV Major Depressive Disorder, any DSM-IV Anxiety Disorder and any DSM-IV Mental Health Disorder.

2.3. Statistical analysis

We used logistic regression to estimate the risk (odds ratio (OR) and 95% confidence intervals (95% CI)) of having a DSM-IV mental health problem by age 21 for each Achenbach measure of A-D and T-P assessed at child age 5 and 14 years. We then examined the sensitivity and specificity of the CBCL (aged 5 and 14 years) and the YSR (aged 14 years) as predictors of CIDI-Auto outcomes at 21 years of age. As both the sensitivity and specificity of a test are influenced by the cut-off point above which

individuals are considered to have problems, we used three different cut-offs – the most extreme 5%, 10% and 20% of scores, to determine if selecting a particular proportion of the sample for early occurring psychopathology may improve prediction or detection of cases at 21 years. At each cut-off, we also calculated the PPV, the proportion (%) of screen-positive individuals who actually did report a mental health problem at age 21, and also the proportion of the population that was identified as positive by a screening tool but was not diagnosed as having mental health problem by DSM-IV (false positives).

Because of resource constraints, only 2563 participants were administered the CIDI-Auto at 21 years (35.5% of the 7223 children originally enrolled in MUSP). Our analyses include only those participants in MUSP for whom we had data at 5 years or 14 years, or both, as well as at 21 years. Thus, sample size varies slightly according to which assessment or screening test is being considered.

To determine whether loss to follow-up at 21 years affected the validity of our findings, we undertook a sensitivity analysis using inverse probability weights reflecting the chances of having missing outcome data (Hogan et al., 2004). We began by constructing a logistic regression model examining the association of all independent variables used in our primary analyses with having complete data or not. The regression coefficients from this model were then used to determine probability weights for the covariates in the main analyses. In the current study, loss to follow-up was predicted by T-P at child age 5 years and A-D at child age 14 years. Children with T-P at 5 years or A-D at 14 years (YSR) were more likely to drop out the study by 21 years (p value <0.05). The results from analyses including inverse probability weighting based on these factors did not differ from the unweighted analyses presented here, suggesting that our results were not substantially affected by selection bias.

3. Results

Of 2563 participants who were administered the CIDI-Auto at the 21-year follow-up, 508 (19.8%) met the DSM-IV criteria for having ever had depression. The corresponding figures for the DSM-IV criteria for a lifetime diagnosis of anxiety were 638 and 24.9%, while for any lifetime DSM-IV mental illness diagnosis, they were 896 and 35.0%.

Table 1 shows that internalizing and T-P at 5 years were modestly associated with anxiety disorder and any mental health disorder at 21 years, respectively. Children who screened positive at 5 years were, at most, 60% more likely to have had any mental health disorder by early adulthood. The associations for both the CBCL and YSR at 14 years with DSM-IV diagnoses were stronger than for the CBCL at 5 years, with YSR scores providing the strongest predictors of DSM-IV mental health problems at 21 years. However, importantly there were no significant differences in the strength of prediction for the various cut-offs examined in the study.

Table 1: Prediction of DSM-IV mental health disorders at 21 years by 5- and 14-year CBCL and YSR

Table 2 presents the relationship between CBCL scores at the 5-year follow-up and CIDI-Auto diagnoses at 21 years. The sensitivity of the screening assessment varied from 5.2% for the T-P score of the CBCL as a predictor of any lifetime mental health disorder, using a cut-off for caseness at screening of 5%, to 28.3% for the internalizing score of the CBCL as a predictor of any DSM-IV diagnosis of anxiety disorder, applying a cut-off of 20% at screening. As expected, specificity fell as

more liberal cut-offs were adopted in pursuit of greater sensitivity. The effect of changes in cut-offs on the PPV, which depends on a combination of specificity and prevalence, was relatively minor. However, the PPVs for all DSM-IV diagnoses were greater than for a DSM-IV diagnosis of depression or anxiety. The proportion of the screened population incorrectly classified as being at high risk (false positives) varied from 55.2% up to 80.8% when different cut-offs were applied to subscales of the CBCL at child age 5 years. Taking the strongest prediction we have as an illustration, there were 566 cases of DSM-IV anxiety at 21 years, of whom 160 (sensitivity of 28.3%) were predicted, leaving 406 cases (71.7% of all cases) at 21 years undetected. Of the 530 persons identified at the 5-year follow-up as being at elevated risk, 370 did not reach the criteria for a case of anxiety at the 21-year follow-up (69.8% false positives).

Table 2: Predicting CIDI-Auto lifetime diagnosis of DSM-IV depression, DSM-IV anxiety and all DSM-IV categories at 21 years using 5-year CBCL screening test scores on internalizing and total problem cores

Endpoint	Sub-scale	Cut-off for cases (%)	Cases at 5 years	Cases at 5 and 21 years	Sensitivity (%) ^a	Specificity (%) ^b	PPV (%) ^c	False positives (%) ^d
Depression (DSM-IV):	Internalizing	5	156	30	6.7	93.2	19.2	80.8
Total N = 2311		10	252	50	11.1	89.1	19.8	80.2
DSM-IV cases = 450		20	529	113	25.1	77.6	21.4	78.6
Anxiety (DSM-IV):	Internalizing	5	157	50	8.8	93.9	31.8	68.2
Total N = 2313		10	253	82	14.5	90.2	32.4	67.6
DSM-IV cases = 566		20	530	160	28.3	78.8	30.2	69.8
All DSM-IV categories:	Total problem	5	98	41	5.2	96.2	41.8	58.2
Total N = 2315		10	201	90	11.3	92.7	44.8	55.2
DSM-IV cases = 796		20	461	182	22.9	81.6	39.5	60.5

^a The proportion of people with DSM-IV diagnosis who had a positive screening test.

^b The proportion of people with a negative DSM-IV diagnosis who had a negative screening test.

^c Positive predictive value: of those who were scored positive on the screening test, the proportion who also subsequently met the DSM-IV criteria for caseness.

^d The proportion of people screening positive who did not meet DSM-IV criteria later.

Table 3 uses the YSR and CBCL data collected at the 14-year follow-up to predict outcomes assessed using the CIDI-Auto at 21 years. There are three patterns in these results that warrant attention. Again, relaxing the criterion for caseness at screening systematically increases sensitivity but reduces specificity. Even with 20% cut-offs at age 14, only minorities of the lifetime cases reported at age 21 are detected. Of the 482 cases of CIDI-Auto depression at the 21-year follow-up, 196 (40.7%) are detected using the 20% cut-off for the YSR. Using the same test and cut-off, there are 636 screen-positive individuals at age 14 years, of whom 440 (69.2%) are false positives.

Table 3: Predicting CIDI-Auto lifetime diagnosis of DSM-IV depression, DSM-IV anxiety and all DSM-IV categories at 21 years using 14 year YSR and CBCL screening test scores on anxiety/depression (A-D) and total problem (T-P) scores

Endpoint	Sub-scale	Cut-off for cases (%)	Cases at 14 years	Cases at 14 and 21 years	Sensitivity (%) ^a	Specificity (%) ^b	PPV (%) ^c	False positives (%) ^d
Depression (DSM-IV): Total N = 2435, DSM-IV cases = 482	YSR (A-D)	5	176	67	13.9	94.4	38.1	61.9
	CBCL (A-D)		173	52	10.8	93.8	30.1	69.9
	YSR (A-D)	10	236	84	17.4	92.2	35.6	64.4
	CBCL (A-D)		217	65	13.5	92.2	30.0	70.0
	YSR (A-D)	20	636	196	40.7	77.5	30.8	69.2
	CBCL (A-D)		595	161	33.4	77.8	27.1	72.9
Anxiety (DSM-IV): Total N = 2438, DSM-IV cases = 598	YSR (A-D)	5	177	81	13.5	94.8	45.8	54.2
	CBCL (A-D)		173	64	10.7	94.1	37.0	63.0
	YSR (A-D)	10	237	99	16.6	92.5	41.8	58.2
	CBCL (A-D)		218	81	13.5	92.6	37.2	62.8
	YSR (A-D)	20	636	239	40.0	78.4	37.6	62.4
	CBCL (A-D)		596	204	34.1	78.6	34.2	65.8
All DSM-IV categories: Total N = 2441, DSM-IV cases = 845	YSR (T-P)	5	124	66	7.8	96.4	53.2	46.8
	CBCL (T-P)		119	53	6.3	95.9	44.5	55.5
	YSR (T-P)	10	240	132	15.6	93.2	55.0	45.0
	CBCL (T-P)		227	110	13.0	92.7	48.5	51.5
	YSR (T-P)	20	513	270	32.0	84.8	52.6	47.4
	CBCL (T-P)		523	231	27.3	81.7	44.2	55.8

^a The proportion of people with DSM-IV diagnosis who had a positive screening test.

^b The proportion of people with a negative DSM-IV diagnosis who had a negative screening test.

^c Positive predictive value: of those who were positive on the screening test, the proportion who also subsequently met the DSM-IV criteria for caseness.

^d The proportion of people screening positive who did not meet DSM-IV criteria later.

Second, although the differences in capacity of the YSR and CBCL to predict outcomes are modest, they consistently favour the YSR, which has slightly better sensitivity and almost identical specificity when compared with the CBCL. Nevertheless, the overall pattern suggests that scores specifically on the A-D subscale of the YSR/CBCL at age 14 provide only a limited prediction of that individual's responses to the CIDI-Auto at 21 years of age. Finally, in the face of only moderate sensitivity and PPVs and imperfect specificity, we again see sizeable proportions of children classified at age 14 as cases who do not go on to report DSM-IV diagnoses at age 21.

4. Discussion

We have used the data from a large prospective cohort to examine the viability of screening children and adolescents to detect psychopathology occurring by early adulthood. Children at ages 5 and 14 years were assessed using the CBCL (5 and 14 years) and the YSR (14 years) at stages in their life course which are arguably sensitive periods and which have been associated with the onset of psychopathology. We have used three different criteria (cut-offs) for caseness at screening at ages 5 and 14 years to predict CIDI-Auto depression, anxiety and all DSM-IV diagnoses at 21 years of age. The findings are consistent in a number of ways. While the sensitivity of the CBCL screening at age 5

improves with the increase in the proportion of individuals accepted as “cases”, even with a 20% cut-off, only about 20% of cases occurring by young adulthood are detected. The disadvantage of relaxing the threshold for “caseness” in childhood is that the vast majority of those selected as likely to manifest pathology subsequently do not meet the DSM-IV criteria for caseness at 21 years.

The results are somewhat more promising when adolescents are screened at 14 years of age. Sensitivity improves with screening in adolescence while specificity remains over 80%. Sensitivity is somewhat better for the YSR compared with the CBCL (self-reports appear to provide better predictions than maternal reports) and the highest cut-offs clearly lead to better prediction of psychopathology at 21 years of age, albeit at the “cost” of more false positives. However, even employing a cut-off on the YSR of 20%, only 32.0% of individuals with any DSM-IV diagnosis by 21 years are detected by the screening assessment. While our data suggest that case detection at or around puberty is preferable to screening in early childhood, the impact on the incidence of subsequent mental ill-health would still be very limited even if a perfectly efficacious intervention were delivered to all screen-positive individuals.

4.1. Clinical implications

The findings from this large scale, population-based prospective longitudinal study raise some important questions about the value of early childhood screening for mental illness. They challenge the case for screening children for depression in early childhood. Our findings favour screening for psychopathology in adolescence rather than early childhood, using child self-reports rather than maternal reports, and using more rather than less liberal criteria for case detection. The main disadvantages of early detection of psychopathology are that the vast majority of persons identified at screening do not go on to become cases at 21 years of age, and the majority of cases at 21 years of age will remain undetected via screening. Substantial numbers of individuals not actually at risk of young adult onset of psychopathology would be incorrectly ‘labelled’ and offered, or subjected to, treatments that are not indicated for them.

Our findings are strongest when the screening test is used to predict any DSM-IV diagnoses at the 21-year follow-up. This suggests that the screening tests we used are non-specific in detecting subsequent cases of impaired mental health, and confirms concerns that the clinical distinction between anxiety and depression is not supported by population data (Moffitt et al., 2007). Judgments about whether to screen for early evidence of mental health problems must depend upon the cost, effectiveness and possible harms associated with any proposed intervention. To be useful, any intervention based on our screening tests would need to be available to a large minority of the population, of relatively low cost/intensity, somewhat effective and with few negative consequences for those who are the subjected to it. Our findings challenge an existing movement encouraging the detection and treatment of those with symptoms of mental illness in early childhood. We have been unable to locate a single paper using a prospective design which disagrees with our data.

4.2. Limitations

There are a number of limitations which must be considered when assessing the above findings. First, we have used a short form of the CBCL (internalizing) subscale at 5 years of age. While this subscale has a high correlation with the full internalizing subscale, it may nevertheless be less discriminating than the full subscale in identifying subsequent caseness. Elsewhere we have demonstrated strong correlations between the short and long forms of the CBCL (Najman et al., 1997).

Second, it could be suggested that the full CBCL should be used at 5 years of age, rather than the internalizing subscale. The internalizing subscale correlated highly with the full scale – it is, after all, a significant component of the full CBCL scale. We have concentrated on the internalizing subscale because we wished to assess its predictive capacity for A-D outcomes at 21 years, these being amongst the commonest mental health problems. More generally, some may argue that we have used the wrong instruments for screening entirely, despite their strong records in the assessment of behaviour in childhood and adolescence, or that age 21 is too soon to expect to see outcomes. While other screening tests are available (e.g. the Strengths and Difficulties Questionnaire – SDQ), comparisons of these tests against each other and external validations suggest that they are strongly correlated and that they produce comparable results ([Goodman and Scott, 1999] and [Jensen et al., 1996]). We do agree that longer follow-up of our cohort is desirable.

A third limitation relates to what is known about the natural history of mental illness. As impairments in mental health may fluctuate over time, many participants in this study may have had mental health problems early in their life course that were not evident at the time of screening. If that were true, it would argue for screening on multiple occasions, or for inclusion of questions about previous mental health problems in screening protocols.

Fourth, only 2563 young adults were administered the CIDI-Auto. While cost considerations were the primary reason for the reduced numbers, there is a possibility that the sample selected may not be more broadly representative of the population. We have found that those lost to follow-up are disproportionately from lower socio-economic backgrounds, and come from families with a more frequent history of mental health problems. However, extensive statistical analyses suggest that those lost to follow-up or not included in aspects of follow-up do not adversely affect the internal validity of the findings from MUSP and may well not undermine their external validity either (Najman et al., 2005). As described in Section 2, we have used inverse probability weighting and found that selective attrition is unlikely to have had any material impact on our results.

5. Conclusions

We have confirmed that while child or adolescent mental health impairment predicts mental health problems in early adulthood, the association is not sufficiently strong to recommend screening and early intervention either in early childhood or even possibly in adolescence. To the extent that there is a case for screening, we found that screening should be delayed until adolescent period, it should involve larger cut-offs than generally used to select caseness in the childhood or adolescent period, and any ensuing intervention would need to satisfy a number of demanding criteria.

The use of screening instruments early in the life course to predict DSM-IV diagnostic outcomes in adulthood involves two somewhat different sources of error. The first is the use of a screening instrument which does not directly map on to diagnostic criteria. While one would expect a screening instrument to be associated with a diagnostic outcome, the association may be of only moderate magnitude. Second, the fluctuating nature of anxiety and depression over the life course may mean that some cases are missed. Irrespective of the study limitations and some possible ambiguity in interpreting the data, the findings are consistent in suggesting that screening children or adolescents for early evidence of impaired mental health may not lead to improved mental health in a population.

Conflict of interest statement

No authors have any actual or potential conflict of interest including financial, personal or other relationships with other people or organizations which could inappropriately influence, or be perceived to influence, our work.

Contributors

J.M. Najman developed the study aims and design. The main analyses were undertaken by J.M. Najman and M.A. Heron. M.A. Heron also drafted the literature review. M.R. Hayatbakhsh and K. Dingle contributed to the data analyses and the interpretation of findings. J.M. Najman and M.A. Heron wrote the first draft of the manuscript. J.M. Najman, M. O'Callaghan, W. Bor and G.M. Williams are responsible for the conceptual development and continued management of the Mater-University Study of Pregnancy and its outcomes and take responsibility for the integrity and accuracy of the data analysis. J.M. Najman and K. Jamrozik edited drafts of the paper and contributed to the discussion and conclusion. All authors contributed to and have approved the final version of the paper.

Role of funding source

Funding for this study was provided by NHMRC Grant 210298. The NHMRC had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Acknowledgments

The authors thank the MUSP participants, the MUSP Research and data collection teams, especially Rosemary Aird, and MUSP Data Manager Greg Shuttlewood for their support. The core study was funded by the National Health and Medical Research Council (NHMRC) of Australia, but the views in this paper are those of the authors and do not necessarily reflect the views of any funding body.

References

- Achenbach, 1991a T.M. Achenbach, Manual for the child behavior checklist/4-18 and 1991 profile, University of Vermont, Department of Psychiatry, Burlington (VT) (1991).
- Achenbach, 1991b T.M. Achenbach, Manual for the youth self-report and 1991 profile, University of Vermont, Department of Psychiatry, Burlington (VT) (1991).
- American Psychiatric Association, 1994 American Psychiatric Association, Diagnostic and statistical manual (4th ed.), American Psychiatric Association, Washington (DC) (1994).
- Birmaher et al., 1996 B. Birmaher, N.D. Ryan, D.E. Williamson, D.A. Brent, J. Kaufman and R.E. Dahl *et al.*, Childhood and adolescent depression: a review of the past 10 years. Part I, *Journal of the American Academy of Child and Adolescent Psychiatry* **35** (1996), pp. 1427-1439.
- Costello et al., 2006 E.J. Costello, A. Erkanli and A. Angold, Is there an epidemic of child or adolescent depression?, *Journal of Child Psychology and Psychiatry and Allied Disciplines* **47** (2006), pp. 1263-1271.

Dierker et al., 2001 L.C. Dierker, A.M. Albano, G.N. Clarke, R.G. Heimberg, P.C. Kendall and K.R. Merikangas *et al.*, Screening for anxiety and depression in early adolescence, *Journal of the American Academy of Child and Adolescent Psychiatry* **40** (2001), pp. 929–936.

Goodman and Scott, 1999 R. Goodman and S. Scott, Comparing the strengths and difficulties questionnaire and the child behavior checklist: is small beautiful?, *Journal of Abnormal Child Psychology* **27** (1999), pp. 17–24.

Hayatbakhsh et al., 2007 M.R. Hayatbakhsh, J.M. Najman, K. Jamrozik, A.A. Mamun, R. Alati and W. Bor, Cannabis and anxiety and depression in young adults: a large prospective study, *Journal of the American Academy of Child and Adolescent Psychiatry* **46** (2007), pp. 408–417.

Hogan et al., 2004 J.W. Hogan, J. Roy and C. Korkontzelou, Handling drop-out in longitudinal studies, *Statistics in Medicine* **23** (2004), pp. 1455–1497.

Jensen et al., 1996 P.S. Jensen, H.K. Watanabe, J.E. Richters, M. Roper, E.D. Hibbs and A.D. Salzberg *et al.*, Scales, diagnoses, and child psychopathology: II. Comparing the CBCL and the DISC against external validators, *Journal of Abnormal Child Psychology* **24** (1996), pp. 151–168.

Kessler et al., 2001 R.C. Kessler, S. Avenevoli and K.R. Merikangas, Mood disorders in children and adolescents: an epidemiologic perspective, *Biological Psychiatry* **49** (2001), pp. 1002–1014.

Kovacs, 1997 M. Kovacs, Depressive disorders in childhood: an impressionistic landscape, *Journal of Child Psychology and Psychiatry and Allied Disciplines* **38** (1997), pp. 287–298.

Lynam et al., 2007 D.R. Lynam, A. Caspi, T.E. Moffitt, R. Loeber and M. Stouthamer-Loeber, Longitudinal evidence that psychopathy scores in early adolescence predict adult psychopathy, *Journal of Abnormal Psychology* **116** (2007), pp. 155–165.

Moffitt et al., 2007 T.E. Moffitt, H. Harrington, A. Caspi, J. Kim-Cohen, D. Goldberg and A.M. Gregory *et al.*, Depression and generalized anxiety disorder: cumulative and sequential comorbidity in a birth cohort followed prospectively to age 32 years, *Archives of General Psychiatry* **64** (2007), pp. 651–660.

Najman et al., 1997 J.M. Najman, B.C. Behrens, M. Andersen, W. Bor, M. O’Callaghan and G.M. Williams, Impact of family type and family quality on child behavior problems: a longitudinal study, *Journal of the American Academy of Child and Adolescent Psychiatry* **36** (1997), pp. 1357–1365.

Najman et al., 2001 J.M. Najman, G.M. Williams, J. Nikles, S. Spence, W. Bor and M. O’Callaghan *et al.*, Bias influencing maternal reports of child behaviour and emotional state, *Social Psychiatry and Psychiatric Epidemiology* **36** (2001), pp. 186–194.

Najman et al., 2005 J.M. Najman, W. Bor, M. O’Callaghan, G.M. Williams, R. Aird and G.J. Shuttlewood, Cohort profile: the Mater-University of Queensland Study of Pregnancy (MUSP), *International Journal of Epidemiology* **34** (2005), pp. 992–997.

Peters and Andrews, 1995 L. Peters and G. Andrews, Procedural validity of the computerized version of the Composite International Diagnostic Interview (CIDI-Auto) in the anxiety disorders, *Psychological Medicine* **25** (1995), pp. 1269–1280.

Peters et al., 1998 L. Peters, D. Clark and F. Carroll, Are computerized interviews equivalent to human interviewers? CIDI-Auto versus CIDI in anxiety and depressive disorders, *Psychological Medicine* **28** (1998), pp. 893–901.

Rutter et al., 2006 M. Rutter, T.E. Moffitt and A. Caspi, Gene–environment interplay and psychopathology: multiple varieties but real effects, *Journal of Child Psychology and Psychiatry and Allied Disciplines* **47** (2006), pp. 226–261

Wilson and Jungner, 1968 J.M.G. Wilson and G. Jungner, Principles and practice of screening for disease, World Health Organization, Geneva (1968).

World Health Organization, 1993 World Health Organization, The ICD-10 classification of mental and behavioural disorders, World Health Organization, Geneva (1993).

World Health Organization, 1997 World Health Organization, 1997. CIDI-Auto version 2.1: Administrator's guide and reference. Sydney: Training and Reference Centre for WHO CIDI.