Young Minds Matter: the second Australian Child and Adolescent Survey of Mental Health and Wellbeing

Confidentialised Unit Record File Technical Manual

Contents

INTRODUCTION

Overview

This technical manual provides information on the release of data from Young Minds Matter: the second Australian Survey of Child and Adolescent Mental Health and Wellbeing. It includes a brief background to the survey, instructions for using the Confidentialised Unit Record File (CURF) microdata, reliability of estimates, contents of the CURF, and the conditions of release. Further information on the YMM survey, including design, methodology, diagnostic tools, scales and measures, data collection, data collection processes, and output files, can be found in the Survey Users Guide included in the CURF documentation.

To protect the privacy of individuals and families who participated in the survey, a small number of variables have been omitted from the CURF. A Detailed Unit Record File (DURF) of the survey data will be maintained by the data custodians, the Telethon Kids Institute. Customized data tables from the DURF will be provided on a cost recovery basis. Requests can be made via email to youngmindsmatter.SMB@telethonkids.org.au

About the survey

Young Minds Matter, a national household survey, was conducted by the Telethon Kids Institute, The University of Western Australia in partnership with Roy Morgan Research on behalf of the Australian Government Department of Health. Fieldwork was conducted from May 2013 through to April 2014. The survey included a face to face diagnostic interview with 6,310 parents and carers of children and young people aged 4-17 years and a self-report questionnaire from 2,967 young people aged 11-17 years.

The survey collected information on the 12-month prevalence of the following mental disorders:

- Anxiety disorders
    - Social phobia
    - Separation anxiety disorder
    - Generalized anxiety disorder
    - Obsessive-compulsive disorder
- Major depressive disorder
- Attention-Deficit/Hyperactivity Disorder
- Conduct disorder

The survey also captured information on the impact of these disorders, health and school services used in their treatment, and overall measures of emotional and behavioural problems. In addition, information on general health, education, family characteristics, basic demographics, and socio-economic characteristics was collected. For the first time in a national Australian survey, young people aged 11-17 years answered questions on symptoms of major depressive disorder, along with their use of services, risk behaviours and contextual information.

YMM is part of the National Survey of Mental Health and Wellbeing Initiative and complements data collected in the population surveys of adults and of people living with psychotic illness conducted in 2007 and 2010 respectively. This is the second time the child and adolescent survey has been conducted, and provides updated national information on prevalence of mental disorders and service use in children and adolescents since the first survey was conducted in 1998.

Further detailed information on survey design and methodology is provided in the Survey User's Guide included in the CURF documentation.

USING THE CURF

About the CURF

The YMM CURF allows for the survey data to be released in a form where survey respondents will not be able to be identified. The aim has been to maximise the usefulness of the data while ensuring individual participants' information remains confidential. Fields that have been removed from the CURF are outlined in the section 'Variables not included on the CURF' on page 12.

Data items

A detailed list of the data items can be found in the metadata spreadsheets downloaded as part of the CURF documentation. The majority of items on the CURF are the same as found on the detailed unit record file. However, where items have been suppressed to protect an individual's confidentiality, this is noted with an 'N' for No under the column 'Released on CURF'. All derived variables listed in the spreadsheets are included on the CURF.

The survey questionnaires were broken down into modules. There were 19 separate modules in the parent or carer questionnaire. In addition there were three modules used to collect information recorded by interviewers during the collection process. There is a separate metadata spreadsheet for each questionnaire module. Additionally there are two spreadsheets describing derived variables created during the data analysis phase. The first lists variables derived by applying the DISC scoring algorithms to information collected in the DISC modules. The final spreadsheet lists all other variables derived from the parent or carer questionnaire.

There were 12 separate modules in the youth questionnaire. There is also one additional spreadsheet listing all variables derived from the youth questionnaire during the analysis phase.

The metadata spreadsheets cover the following modules:

Parent or carer interview:

01. Household Record Form
02. Education Parent
03. Child Health Parent
04. SDQ Parent

05. DISC Intro Parent
06. DISC Social Phobia Parent
07. DISC Separation Anxiety Parent
08. DISC Generalised Anxiety Parent
09. DISC Obsessive Compulsive Parent
10. DISC Major Depressive Parent
11. DISC ADHD Parent
12. DISC Conduct Parent
13. DISC Oppositional Problem Behaviours Parent
14. Impact on Functioning Parent
15. DISC Main Condition Parent
16. Bullying Parent
17. Service Use Parent
18. Bullying Parent
19. Family Information Parent
20. Blocklisting Assessment Form
21. SA1 Assessment
22. Interviewer Impressions
23. DISC Derived Variables
24. Other Derived Variables

Youth Self-report:

01. Education Youth
02. Internet Youth
03. SDQ Youth
04. CHU-9D Youth
05. K10 Youth
06. MDD Youth
07. Service Use Youth
08. Family Information Youth
09. Risk Behaviours Youth
10. Psychosis Screener Youth
11. Self Esteem Youth
12. Protective Factors Youth
13. Derived Variables Youth

Structure of the spreadsheets

The metadata spreadsheets include the full text of each question as included in the computer assisted interviewing software used in the survey. Where there was introductory remarks made at the beginning of a set of questions this is also included in the spreadsheets, but does not correspond with a variable on the data set. The spreadsheets also indicate the allowable response codes, and describe the values that correspond to each of these codes. Most questions in the survey were

yes/no questions. Where other responses were allowed, the list of response options was read to the respondent by the interviewer, or the respondent was shown a list of response options on a show card. Where show cards were used this is indicated in the question text. In addition to the allowable response values, for each question in the survey (with the exception of questions in the Strengths and Difficulties Questionnaire) there were separate codes used to indicate if the respondent either did not know the answer to the question or if the respondent refused to answer the question. Where a respondent was not asked a question because of the sequencing of the questionnaire, the variable is left blank.

Derived variables

The metadata spreadsheets included in the CURF documentation outline all of the derived variables created during data analysis and included on the CURF. To assist users to reproduce tables similar to the ones published on the YMM website and in the main report, listed below are the main variables used in the final analysis and presentation of survey data.

Diagnostic status with impairment criteria applied

Responses to the DISC questions were scored using algorithms provided by the DISC custodians at Columbia University in New York. The variables have been named using the standard naming conventions used in the DISC scoring algorithms. The DISC algorithms provide a number of variables for each disorder which indicate if the child or adolescent met the diagnostic criteria for the disorder, the number of diagnostic criteria met, the duration of the disorder, and the age of first onset. Variables are named using the following conventions:

The first letter of the variable name is either 'P' or 'Y' indicating whether the variable is based on parent or carer report ('P') or youth self-report ('Y'). The second and third letters indicate the disorder module as follows:

    so - Social phobia
    sa - Separation anxiety
    ga - Generalised anxiety
    oc - Obsessive-compulsive disorder
    md - Major depressive disorder
    ad - Attention-Deficit/Hyperactivity disorder
    cd - Conduct disorder
    od - Oppositional problem behaviours

The fourth letter indicates the time period - 'm' signifies met criteria in the last four weeks, and 'y' signifies met criteria in the last 12 months.

The indicators of diagnostic status used in the main report were based on diagnostic status in the last 12 months, with impairment criteria applied as follows:

    psoyim – Social phobia
    psayim – Separation anxiety
    pgayim – Generalised anxiety

pocyim – Obsessive-compulsive disorder

pmdyim – Major depressive disorder

padyim – ADHD

pcdyim – Conduct disorder

panxyim – Any anxiety disorder

panxdepyim – Any anxiety or depressive disorder

pbehyim – Any behavioural disorder (ADHD or CD)

pDISCyim – Any DISC disorder, excluding oppositional problem behaviours

podyim – Oppositional problem behaviours

ymdyim – Major depressive disorder – youth report

Impact on functioning

An impact on functioning module that expanded on the DISC impairment questions was developed by the survey team to assess the impact of symptoms of mental disorders on four life domains (school or work, family, friends, and self). Several methods were employed in developing the scoring of the questions. The derived variables for each are included in the CURF. Further detail on the development of the impact on functioning scale can be found in the Survey User's Guide and Measuring Severity Technical Report included in the CURF documentation.

Pseverity – conceptual mapping

pseverity_IRT – Item response theory

pseverity_IRTP – Item response theory pooled analysis

The item response theory pooled analysis approach has been used to describe impact on functioning in the main survey report.

Principal condition

If a child or adolescent had symptoms of more than one mental disorder, the parent or carer was asked which symptoms bothered or upset the child or adolescent the most. This was used to assign a principal condition to each child or adolescent with more than one mental disorder. There are several ways in which this can be derived, and there are multiple variables on the CURF representing these different approaches. During the interview, the computer assisted interview identified the number of mental disorder modules where the child or adolescent reached a sub-threshold level, that is where half or more of the number of symptoms required for diagnostic status were reported. This calculation did not take into account impairment of functioning, which had not been assessed at that stage of the questionnaire. Because of this methodology, the parent or carer might have been asked to assign a principal condition among disorders where the child or adolescent did not meet full diagnostic criteria with impairment included in the assessment. In addition, oppositional problem behaviours was included in the list of conditions for the principal condition question if the child or adolescent met the sub-threshold criterion, although oppositional problem behaviours was not included in the overall measure of mental disorders in the main survey report.

Where the parent or carer nominated as principal condition a condition for which the child or adolescent did not meet the full diagnostic criteria three approaches were taken to assigning a principal condition:

a) use the condition nominated by the parent or carer if the child or adolescent met diagnostic criteria for that condition, otherwise treat it as not stated

b) use the condition nominated by the parent or carer, regardless of whether the child or adolescent met diagnostic criteria for that condition or not

c) use the condition nominated by the parent or carer if the child or adolescent met diagnostic criteria for that condition, otherwise choose a condition for which the child or adolescent met full diagnostic criteria using an algorithm based on severity of each condition. The following algorithm was applied:

- If the parent or carer has nominated Social Phobia and the child or adolescent meets full diagnostic criteria for Separation Anxiety, Separation Anxiety.

- If the parent or carer has nominated Separation anxiety, and the child or adolescent meets full diagnostic criteria for Social Phobia, Social Phobia.

- If the parent or carer has nominated Social Phobia, Separation Anxiety or Obsessive Compulsive Disorder, and the child or adolescent meets full diagnostic criteria for Generalised Anxiety Disorder, Generalized Anxiety Disorder.

- If the impact score for ADHD is higher than the impact score for either conduct or anxiety/depression, ADHD.

- If the impact score for conduct is higher than the impact score for ADHD or anxiety/depression, and the child or adolescent meets full diagnostic criteria for Conduct Disorder, Conduct Disorder.

- If the impact score for conduct is higher than the impact score for ADHD or anxiety/depression, and the child or adolescent meets full diagnostic criteria for Conduct Disorder, Conduct Disorder.

- If the impact score for conduct is higher than the impact score for ADHD or anxiety/depression, and the child or adolescent meets full diagnostic criteria for Oppositional Problem Behaviours, Oppositional Problem Behaviours.

- Otherwise based on the following hierarchy:

    - Major Depressive Disorder
    - Generalised Anxiety Disorder
    - Social Phobia
    - Separation Anxiety
    - Obsessive-Compulsive Disorder
    - Conduct Disorder

- ADHD
- Oppositional Problem Behaviours

Principal condition variable	Approach	Includes Oppositional problem behaviours as a permitted principal condition	Diagnostic status determined including impairment criteria
MainDisorder	a	Y	N
MainDisorder1	b	Y	N
MainDisorder2	c	Y	N
MainDisorderB	a	N	N
MainDisorderB1	b	N	N
MainDisorderB2	c	N	N
MainDisorderIM	a	Y	Y
MainDisorder1IM	b	Y	Y
MainDisorder2IM	c	Y	Y
MainDisorderBIM	a	N	Y
MainDisorderB1IM	b	N	Y
MainDisorderB2IM	c	N	Y

Variable *MainDisorderBIM* was used to identify principal condition in the main survey report.

CURF file structure

The CURF contains two files, a parent or carer file with data captured from the parent or carer interview and a youth file that contains data from the youth self-report questionnaire. Each household has been assigned at random a unique household identifier (HouseholdID). This enables the youth and parent data from the same household to be linked for analysis.

Use of weights

Due to the complex nature of the YMM sample design, three sets of weights have been derived. Data from the parent and carer questionnaire have been weighted to represent all children aged 4-17 years in Australia, and also to represent all families with one or more children aged 4-17 years in Australia. Both of these weights are on the parent file. The youth self-report data have been separately weighted to represent all 11-17 year-olds in Australia, based on the response rates to the youth self-report questionnaire.

Three weight variables have been provided on the CURF:

- The population of all children aged 4-17 years in Australia (ChildWeightC)
- The population of all families with one or more children aged 4-17 years in Australia (FamilyWeightC)
- The population of all adolescents aged 11-17 years in Australia (YouthWeightC)

It is expected that most of the analyses from YMM data will be based on the population of all children, but the creation of family level weights also allows for analyses to be conducted at the family level, where appropriate.

The survey sample was selected in a multi-stage design. First geographic areas (SA1s) were selected, then households, and then children within selected households. Both the survey weights and the clustered nature of the sample design should be taken into account when calculating estimates from the survey and measures of precision such as confidence intervals. To facilitate this, a cluster identifier variable, named CLUSTER, has been included on the CURF files.

An example of how to calculate weighted estimates of proportions and confidence intervals from the survey data using SAS is shown below:

```
proc surveyfreq data = CURF.parent;
    tables sex * pdiscyim / row cl clwt;
    cluster cluster;
    weight ChildWeightC;
run;


proc surveyfreq data = CURF.youth;
    tables sex * ymdyim / row cl clwt;
    cluster cluster;
    weight YouthWeightC;
run;
```

Record identifiers

Each household has been assigned at random a unique identifier (HouseholdID). Where applicable this allows parent or carer data to be linked to the relevant young person self-report.

Continuous data items

When analysing continuous data items it is important to exclude the special codes that have been assigned for when data was unable to be collected (e.g. 'don't know' or 'refuse to answer'). For example, age of household members has allowable values of 1 through to 110 years and the following special codes:

999 – Don't know
777 – Refuse to answer

These codes are included in the metadata spread sheets as part of the CURF documentation. Care needs to be taken to exclude these values from analysis when categorising higher values for ranges and calculation of means, medians and other summary statistics.

Income items

Equivalised household income is included on the CURF. It is calculated by using an equivalence factor based on the 'Modified OECD' equivalence scale and then dividing total household income by that

factor. The factor is calculated using a weighted household size, as follows: The first adult contributes 1 point, any additional adults or young people aged 14 or over contribute 0.5 points, and any children under the age of 14 contribute 0.3 points.

This allows comparisons of income between households of different sizes to be made.

Household income and equivalised household income are provided in both approximate quintiles and terciles. The quintiles used were:

Household income before tax:

>   Less than $41,599 per year (Less than $799 per week)
> $41,600-$77,999 per year ($800-$1,499 per week)
> $78,000-$103,999 per year ($1,500-$1,999 per week)
> $104,000-$155,999 per year ($2,000-$2,999 per week)
> $156,000 or more per year ($3,000 or more per week)

Equivalised household income:

>   Less than $20,000 per year
> $20,000-$32,999
> $33,000-$44,999
> $45,000-$66,999
> $67,000 or more per year

The terciles used were:

Household income before tax:

>   Less than $52,000 per year (Less than $1,000 per week)
> $52,000-$129,999 per year ($1,000-$2,499 per week)
> $130,000 or more per year ($2,500 or more per week)

Equivalised household income:

>   Less than $28,000 per year
> $28,000-$49,999 per year
> $50,000 or more per year

Multiple response items

A number of questions included in both the parent and youth interview had multiple response options. This means that when answering the question respondents could select one or more options as their response. For example, parents were asked to identify from a list if their child had any common chronic health conditions for which they could select more than one health condition.

On the CURF each response for the multiple response questions is treated as a separate data item with a yes or no response option for each item.

Geographic characteristics

The CURF contains two geographic classification variables. Firstly, households have been classified in to Greater capital cities and Rest of state (MetRur). In addition, the ABS remoteness index, Accessibility/Remoteness Index of Australia (ARIA+), has been applied (Remoteness).The one percent most remote SA1s were excluded from the sampling frame, therefore, the two ARIA classification levels "Remote Australia" and "Very Remote Australia" have been combined into a single level on the CURF. Both of these classifications are based on the Australian Statistical Geography Standard (ASGS) as developed by the ABS.

If any further geographical classification is required for analysis, a special request can be made to the data custodians at the Telethon Kids Institute for this to be undertaken on the detailed unit record file. This service will be provided on a cost recovery basis.

Reliability of estimates

Two types of errors are possible when calculating estimates from a survey sample such as Young Minds Matter; sampling error and non-sampling error.

Sampling error

Sampling error is the difference that may occur by chance between an estimate derived from a sample and the 'true value' (or full population census value) that would be obtained if the whole population were included. For example, the YMM sample comprised 6,310 families with children aged between 4-17 years, however, according to the 2011 Census of Population and Housing, there were over two million families with children and young people aged 4-17 years in Australia in 2011. Sampling error can be reliably measured from the sample itself, and can be described through the use of measures such as standard errors or confidence intervals which may be used to indicate how much an estimate might have varied by chance because only a sample of the full population was interviewed.

Non-sampling error

Non sampling errors can occur in any data collection, regardless of whether it is a sample survey or a full enumeration, at any stage during the survey process. These may include respondents misunderstanding or not answering questions, responses being incorrectly entered or errors in processing the survey data.

Where possible, every effort was made to minimise non sampling error in the survey. Questionnaires were carefully designed and tested for comprehension. Interviewers all completed a three day training program and the use of computer tablets for administration of the questionnaires helped to reduce data entry errors. Data was also comprehensively checked and edited post survey. Further detail on editing on the survey data is outlined in the Survey User's Guide included in the CURF documentation.

Response rate

While extensive efforts were made to maximise the response rate to YMM, the final response rate was 55%. The representativeness of the responding sample was assessed by comparing the YMM

survey sample to 2011 Census data. The survey included some instruments such as the SDQ and K10, which are included in other collections, so comparisons were also made to the 1999 and 2004 British Child and Adolescent Mental Health Survey (BCAMHS) and the 2007-2008 National Health Survey. As a result of this analysis, adjustments were made to the weighting of the data. Further detail on representativeness of the sample can be found in the Survey User's Guide included in the CURF documentation.

Variables not included on the CURF

To protect the confidentiality of participants in the survey, a small number of variables on the detailed unit record file have been removed in preparation of the CURF. These include:

- Names and addresses of survey participants

- Date of birth of the study child (age is provided on the CURF)

- Country of birth (other than the 12 most common countries of birth)

- State of birth and residence

- Statistical Area 1 (SA1) (geographic area) has been replaced with a cluster number

- Aboriginal or Torres Strait Islander status

- Type of school child is attending

- Other chronic health condition (write in item) - these have been coded to ICD10 chapter

- If child has recently been bereaved, name and date of birth of person who died

- Interview date

- Names of any medications taken by survey child

Contents of the CURF

The CURF is available from the Australian Data Archive in a number of formats. These include SAS, SPSS and STATA data sets.

As the analysis of survey data was completed in SAS 9.4 for those variables that relate to specific questions the question text is supplied in the 'Label' field on the dataset. In transferring the CURF to other statistical programs, such as SPSS and STATA, limits on the length of labels have meant that at times the full question text cannot be displayed in the dataset. To enable CURF users to refer back to specific questions in the survey questionnaires (also supplied in the CURF documentation) the table in Appendix A provides a variable name 'cheat sheet'.

Supporting documentation

The following documentation is provided with the CURF:

- Library of formats.

- Survey User's Guide – A comprehensive guide that describes the survey background and design, survey methodology, diagnostic tools, scales and measures, data collection processes, and output files.

- Metadata spread sheets – include variable names, question text, allowable response values, formats and labels for each module in the parent interview and youth self-report questionnaire.

- Survey Questionnaires – copies of the parent interview and youth self-report questionnaire.

- Consent booklets and survey brochures – copies of the consent booklets and parent and youth survey brochures.

- Measuring severity technical report – detailed report on the development of the impact on functioning questions by the survey team.

- Young Minds Matter Technical Report – A report prepared by Roy Morgan Research detailing the survey fieldwork.

Survey products

In addition to the CURF, the following survey products have been produced:

The Mental Health of Children and Adolescents: Report on the Second Australian Child and Adolescent Survey of Mental Health and Wellbeing. Available on the Department of Health website (www.health.gov.au).

The survey website (www.youngmindsmatter.org.au) also provides access to some basic analysis of the survey data via a Survey Results Query Tool. This facility allows users to create downloadable Excel-based tables from menus of pre-determined variables.

The Detailed Unit Record File (DURF) will be maintained by the Telethon Kids Institute. Customised analysis of the DURF can be requested and will be provided on a cost recovery basis. Requests can be made via: youngmindsmatter.SMB@telethonkids.org.au

Series of peer reviewed papers to be published in scientific journals e.g. Australian and New Zealand Journal of Psychiatry.

Conditions of release

Release of the CURF

The CURF has special restricted access condition which requires all researchers to have obtained ethics approval for their research from their own relevant ethics body before seeking access to the YMM CURF. Use of the CURF is only for the statistical purposes as specified by the researcher in the required ethics application.

By agreeing to the access conditions for this CURF, researchers will be providing an undertaking to maintain respondent confidentiality by using the CURF for statistical purposes only and not attempting to identify any particular persons or to attempt to match the YMM CURF information with any other lists or data sets of persons or in any other way.

APPENDIX A

The analysis of survey data was completed in SAS 9.4. For those variables that relate to specific questions the question text is supplied in the 'Label' field on the dataset. In transferring the CURF to other statistical programs, such as SPSS and STATA, limits on the length of labels have meant that at times the full question text cannot be displayed in the dataset. To enable CURF users to refer back to specific questions in the survey questionnaires (also supplied in the CURF documentation) the table below provides a variable name 'cheat sheet'.

The first two to three letters of the variable name indicate which module the variable comes from, the numbers and letters following this refer to the question number within the module. So, for example, pch2a refers to question 2A in the parent child health module.

Variable name	Module
hrf	Household Record Form
ped	Parent Education
PCH	Parent Child Health
p1c	Parent SDQ 4-10 years
p2c	Parent SDQ 11-17 years
int	Parent DISC-IV Introduction
PSO	Parent DISC-IV Social phobia
PSA	Parent DISC-IV Separation anxiety disorder
PGA	Parent DISC-IV Generalised anxiety disorder
POC	Parent DISC-IV Obsessive-compulsive disorder
PMD	Parent DISC-IV Major depressive disorder
asv	Parent Impact on function questions anxiety/major depressive disorder
PAD	Parent DISC-IV Attention-deficit/Hyperactivity disorder (ADHD)
hsv	Parent Impact on function questions ADHD
PCD	Parent DISC-IV Conduct disorder
POD	Parent DSIC-V Oppositional problem behaviours
csv	Parent Impact on function questions conduct disorder
mp	Parent main problem
pb	Parent bullying
SUP	Parent service use
PFI	Parent family information
pcd	Parent demographics - primary carer
scd	Parent demographics - secondary carer
hhd	Parent demographics – household
iip	Interviewer impressions

SAA	SA1 assessment form
yed	Youth education
yiu	Youth internet use
p3c	Youth SDQ 11-17 years
chu	Youth child health utility scale 9D
ych	Youth Kessler 10+
yint	Youth DISC-IV introduction
ymd	Youth DSIC-IV major depressive disorder
yds	Youth Impact on functioning major depressive disorder
suy	Youth Service use
yfi	Youth family information
yrb	Youth risk behaviours
ych	Youth psychosis screener
yse	Youth self esteem
ypf	Youth protective factors