Survey Research Methods (2018)
Vol. 12, No. 2, pp. 161-176
doi:10.18148/srm/2018.v12i2.7225

ISSN 1864-3361
http://www.surveymethods.org

Respondent mental health, mental disorders and survey interview outcomes

Francisco Perales
Institute for Social Science Research
The University of Queensland
Brisbane, Australia

Bernard Baffour
School of Demography
The Australian National University
Canberra, Australia

Poor mental health and mental disorders are amongst the largest causes of disease burden across the globe, and in developed countries mental illness is on the rise. Studies of the predictors and consequences of ill mental health and mental disorders often rely on surveys. However, there is very little evidence of whether or not there are differences in the ways in which individuals with good and poor mental health and with and without mental disorders engage with the survey interview process, and on their subsequent survey interview outcomes. We examine the associations between respondent mental health, mental disorders and survey interview outcomes using 14 years (2001–2014) of annual, nationally-representative, Australian panel data ($n \approx 200,000$) and state-of-the-art multilevel regression models. We find that individuals with poorer mental health and mental disorders are generally more likely than individuals with better mental health and without mental disorders to be deemed by interviewers as being suspicious of the study, experiencing issues understanding survey questions, and being uncooperative. We also find that these individuals are comparatively more likely to experience panel attrition, complete interviews featuring higher item-level missing data, and fail to complete/return self-complete questionnaires. While the magnitude of these effects is moderate, our findings suggest that data collectors, researchers and policymakers need to remain cognizant of potential issues emerging from differences in the ways in which individuals with poorer and better mental health and without and without mental disorders engage in social surveys.

Keywords: paradata; interviewer observations; panel attrition; missing data; multilevel models; survey methods

1 Introduction

Poor mental health and mental disorders are amongst the largest causes of disease burden at a global level (World Health Organization., 2004; World Health Organization., 2008). As a result, understanding their predictors and consequences is a fundamental goal of health research, policy and practice, and has been the focus of a wealth of interdisciplinary research. Many contemporary studies of mental health and mental disorders rely on the analysis of population surveys. For example, a Scopus search for research articles published in 2016 in which the terms “mental health” and “survey” appear in the article title, abstract or keywords yields 2,595 items.

Within this context, survey collectors, methodologists and researchers should be concerned with how individuals with poor mental health or who experience mental disorders engage with the survey interview process. This is an important exercise, as there are reasons to expect comparatively poorer survey interview outcomes amongst such individuals. Based on information processing theory (Schwarz, 2007), these may include higher-than-average levels of discomfort when engaging in the social interactions involved in a survey interview, relatively lower interest and motivation in answering the survey questions, and reduced faculties in cognitive capabilities which are important for the processing of survey questions.

Despite this, research comparing the survey interview outcomes of individuals with good and poor mental health and with and without mental disorders is surprisingly scarce. We fill this gap in knowledge by comparing interviewer ratings of respondent engagement with the survey interview (IRRESI) and indicators of objective survey interview outcomes (OSIO) between respondents with poorer and better mental health, and with and without mental disorders. We consider three indicators of IRRESI (interviewer reports of survey respondents being suspicious of the study, having issues understanding the survey questions, and being uncooperative), and

Contact information: Francisco Perales, Institute for Social Science Research, The University of Queensland, Long Pocket Precinct, 80 Meiers Rd, Building C, Indooroopilly, Brisbane QLD 4068 (f.perales@uq.edu.au)

three indicators of OSIO (panel attrition, item-level missing data, and failure of return a self-complete questionnaire). Our analyses use 14 years of annual, nationally representative, panel data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey (Summerfield et al., 2015; The Melbourne Institute of Applied Economic and Social Research (MIAESR)., 2014), and state-of-the-art, three-level, cross-classified models.

2 Background

Information processing theory perspectives in survey methodology pose that when respondents are presented with a question in the context of a survey interview they engage in a series of mental processes before formulating an answer (Schwarz, 2007). The dominant approach comprises a four-phase model of survey response: question interpretation (i.e. how the respondent understands the interviewer request, Phase 1), information retrieval (i.e. the process of recalling the necessary information asked about, Phase 2), judgement (i.e. deciding which of the retrieved information will be shared with the interviewer, Phase 3), and response editing (i.e. formulating a response in actual words, Phase 4) (Tourangeau, Rips, & Rasinski, 2000, p. 8). It has been argued that socio-demographic factors, such as age, cultural background or physical health can affect how survey respondents engage in each of these phases by influencing individuals' capabilities and schemata (Esposito & Jobe, 1991; Groves, 2004; Groves et al., 2009; Tourangeau et al., 2000).

Drawing on the information processing framework, we argue that symptoms associated with poor mental health or mental disorders have the potential to alter the survey response process in ways that result in suboptimal survey interview outcomes. Following the literature, we conceptualize mental health and mental disorders as "points in a continuum" (see Satcher, 2000). Mental health is "the successful performance of mental function, resulting in productive activities, fulfilling relationships with other people, and the ability to adapt to change and to cope with adversity" (Satcher, 2000, p. 6). The absence of psychological distress, characterized by anxiety and depression, contributes to good mental health (Kessler et al., 2002). In contrast, mental disorders are "health conditions that are characterized by alterations in thinking, mood, or behavior (or some combination thereof) associated with distress and/or impaired functioning" (Satcher, 2000, p. 7). These are more severe and specific conditions than poor mental health or high psychological distress. As detailed in the latest edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V American Psychiatric Association., 2013), these include – amongst others – depressive disorders, anxiety disorders, neuro-cognitive disorders, or intellectual developmental disorders. Mental health and mental disorders should thus not be equated, yet in this paper we are interested in the impacts of both of these constructs on survey outcomes. While there is variation in the nature and severity of mental health issues and disorders, we identify three general mechanisms which could produce these associations and which relate to motivational as well as cognitive processing.

First, the very nature of some mental health problems may lead individuals to experience higher-than-average levels of discomfort when engaging in certain types of social interactions. This discomfort could result in issues affecting primarily the judgement phase of the four-phase model of survey response outlined before, by altering the degree of openness and honesty of survey respondents. For example, individuals suffering from neurotic disorders, such as anxiety disorders and social phobias, display heightened fear of being criticised or embarrassed in everyday situations, particularly when interacting with strangers and when operating within unfamiliar settings or situations. This applies strongly to the context of face-to-face survey interviews, in which respondents are asked multiple personal questions by a stranger over a prolonged period of time following a highly structured and rigid communication mode (Perales, Baffour, & Mitrou, 2015). In addition, there is a social stigma against people who have poor mental health or mental disorders (Link, Phelan, Bresnahan, Stueve, & Pescosolido, 1999; World Health Organization., 2010), which may make these individuals less open to fully engage in survey interviews due to perceived discrimination and power imbalances. In this regard, feeling "down in the dumps", "worthless", "nervous" or "hopeless" are part of general questionnaires used to assess mental health and psychological distress (as explained below). This suggests that, when faced with such an unfamiliar situation, individuals with these symptoms may be more likely to be apprehensive or mistrustful of interviewers, less likely to ask clarification questions about the meaning of survey items, and less willing to provide open, accurate and truthful answers to survey items. It also suggests that such symptoms may interfere with survey interviewers' ability to establish rapport with these respondents. In both cases, the end result is likely to be survey interviews characterized by imprecisions, suspicions and uncooperativeness, and a low propensity to remain engaged in prospective surveys (Watson & Wooden, 2009). While there is no empirical evidence on these propositions, these arguments resonate with findings from studies in cognate fields or inquiry. For example, research documents challenges faced by clinical staff in establishing successful interpersonal communication and cooperation strategies with hospital patients with mental health issues (Eren & Şahin, 2016; Treloar, 2009).

Second, poor mental health may also lead to lower interest and motivation when participating in a survey interview (or motivational processing). This is important, as engaged and enthusiastic respondents are pivotal in increasing the quality of the information generated from survey participants (Groves et al., 2009). Comparatively low respondent engagement and motivation levels could result in issues affecting the judgement phase of the aforementioned model of survey response (e.g. deciding not to share certain details with the interviewer), but also other phases (e.g. if respondents are less willing to expend effort paying attention to questions, retrieving the information necessary to answer, or editing their responses). For example, individuals who suffer mood (affective) disorders usually display symptoms characterized by depression, apathy or anhedonia, can have comparatively low energy and high fatigue, reduced problem-solving capabilities, and a reduced ability to concentrate. These symptoms are also dimensions of general measures of mental health and psychological distress, such as those used in our empirical analyses (details below). As a result, respondent burden might be comparatively higher for these individuals when presented with the same survey interview, which would negatively impact respondent effort, response errors and willingness to continue participating in the survey. Low energy and high fatigue (which are items falling within the spectrum of psychological distress) may translate into lower capabilities to focus on the task, and maintain attention, concentration and motivation over the duration of the survey interaction, particularly if the interview is long. Similarly, negative emotional states (or moods) can lead respondents to spend comparatively little cognitive effort in answering questions, or satisficing (Krosnick, 1991). This would apply to individuals with depressive symptoms not only due to the general emotional symptoms associated with poor mental health levels or the experience of a specific mental disorder, but also if they are more prone to have negative moods elicited by virtue of participating in the survey. This could occur if respondents' moods become more negative by, for example, being presented with unexpected questions or questions perceived to be intrusive, or facing unfamiliar interviewer behaviours that make them uncomfortable (Esposito & Jobe, 1991). These propositions apply particularly strongly to people suffering from personality disorders such as bipolar disorders, whose condition is defined by the experience of sudden mood changes. Depression and anhedonia – which are present in both poor general mental health levels and specific mental disorders – are also characterized by an inability to perceive intrinsic value in undertaking routine and non-routine activities, or to derive pleasure from social and civic activities. This is important, as most surveys are imbalanced social exchanges from which respondents obtain (relatively) small direct gains. Hence, individuals with these symptoms are likely to perceive lower intrinsic rewards in undertaking the cognitive processes necessary to provide accurate survey answers, e.g. information retrieval and assessment. Taken together, these arguments suggest that survey interviews involving individuals with poor mental health and/or certain mental disorders may be characterized by comparatively low levels of engagement and cooperation, and higher-than-average chances of incomplete or discontinued participation.

Third, some mental disorders are comorbid with reduced faculties in cognitive capabilities important for the cognitive processing required for the successful completion of face-to-face survey interviews. This includes capabilities such as the ability to concentrate, abstract thinking, memory retention, or mathematical computation (Koenen et al., 2009). Low cognitive functioning could affect several phases of the survey response model: the interpretation phase (e.g. via poor question comprehension), the information retrieval phase (e.g. via inability to recall/remember information), and the response editing phase (e.g. via inability to formulate an appropriately formatted answer). Some mental disorders captured in the DSM-V are defined in terms of the cognitive difficulties that they entail, e.g. dyslexia, attention-deficit disorders or intellectual disability. Others, such as depression, involve temporary cognitive dysfunctions. Dementia – one of the most prevalent umbrella mental disorder in elderly populations in developed countries (World Health Organization., 2017) – is also characterized by the impairment of cognitive, language, memory, perception and personality functioning. As a result, individuals with these symptoms may on average experience more issues understanding and responding to the survey questions. This is consistent with evidence indicating that recall bias amongst people suffering from depression substantially affects survey estimates (Kruijshaar et al., 2005; Patten, 2003), and that poor cognitive ability is related to difficulties answering survey questions and suboptimal survey responses, such as acquiescence (see e.g. Borgers, de Leeuw, & Hox, 2000; Meisenberg & Williams, 2008; Sigelman et al., 1980). The cognitive capacity channel potentially connecting mental wellbeing and survey responses should relate almost exclusively to mental disorders rather than poor mental health levels more broadly.

Collectively, these general principles lead us to hypothesize that poor mental health and mental disorders will be associated with lower IRRESI and OSIO. Nevertheless, to our knowledge, no previous empirical studies have examined these associations.

3 Data and methods

3.1 Dataset

We examine the associations between mental health, mental disorders and survey outcomes using data from the HILDA Survey (Watson & Wooden, 2012). This is a household panel study conducted by the Melbourne Institute of Applied Economics and Social Research at the University of Melbourne, and which collects annual information from

Table 1
Sample descriptive statistics

	Observations	Mean	Std. Dev.	Min	Max
Outcome variables					
Interviewer assessment: Respondent was suspicious of the study after the interview	200, 237	0.02	-	0	1
Interviewer assessment: Respondent had issues understanding the survey questions	200, 238	0.04	-	0	1
Interviewer assessment: Respondent was not cooperative during the interview	200, 239	0.02	-	0	1
Survey outcome: Panel attrition in the next survey wave	182, 799	0.07	-	0	1
Survey outcome: At least0.02 of responses to survey questions are missing	200, 308	0.10	-	0	1
Survey outcome: Respondent did not complete/return the SCQ	200, 311	0.10	-	0	1
Key explanatory variables					
SF-36 Mental Health Inventory	178, 252	74.20	17.14	0	100
K10 Psychological Distress Scale	53, 238	15.72	6.30	10	50
Respondent has a mental illness requiring help/supervision	173, 301	0.02	-	0	1
Respondent has difficulty learning/understanding things	173, 301	0.01	-	0	1
Respondent has a nervous/emotional condition requiring treatment	173, 301	0.04	-	0	1
Control variables					
Female	200, 311	0.53	-	0	1
Age in years	200, 311	44.00	18.56	15	101
Partnered	200, 197	0.62	-	0	1
Number of adults in the household	200, 311	2.30	1.04	1	9
Number of children in the household	200, 311	0.59	1.00	0	11
Ethno-migrant background	200, 260				
Australian born, not Indigenous	-	0.76	-	0	1
Australian born, Indigenous	-	0.02	-	0	1
Migrant from English-speaking background	-	0.10	-	0	1
Migrant from non-English-speaking background	-	0.12	-	0	1
Highest educational qualification	200, 201				
Degree or higher degree	-	0.21	-	0	1
Professional qualification	-	0.28	-	0	1
School year 12	-	0.15	-	0	1
Below school year 12	-	0.35	-	0	1
Employment status	200, 311				

Continues on next page

Continues from previous page

	Observations	Mean	Std. Dev.	Min	Max
Employed (including self-employment)	-	0.63	-	0	1
Not in the labour force	-	0.33	-	0	1
Unemployed	-	0.04	-	0	1
Annual household disposable income (in $10,000s)	200, 311	8.59	6.47	0	201
Area remoteness	200, 311				
Major city	-	0.62	-	0	1
Inner regional area	-	0.24	-	0	1
Outer regional, remote or very remote area	-	0.13	-	0	1
Socio-Economic Index for Areas	200, 265				
1st quintile	-	0.20	-	0	1
2nd quintile	-	0.20	-	0	1
3rd quintile	-	0.20	-	0	1
4th quintile	-	0.20	-	0	1
5th quintile	-	0.20	-	0	1
State of residence	200, 311				
New South Wales	-	0.30	-	0	1
Victoria	-	0.25	-	0	1
Queensland	-	0.21	-	0	1
South Australia	-	0.09	-	0	1
Western Australia	-	0.09	-	0	1
Tasmania	-	0.03	-	0	1
Northern Territory	-	0.01	-	0	1
Australian Capital Territory	-	0.02	-	0	1
Number of times previously interviewed	200, 311	5.81	3.91	1	14
First contact with interviewer	200, 311	0.51	-	0	1
Interviewer workload	200, 311	123.78	57.61	1	389
Survey year	200, 311	2008	4.13	2001	2014

HILDA Survey data, Australia, 2001–2014.

the same respondents over the 2001–2014 period. It is one of the largest and best-known panel surveys in the developed world and part of the Cross National Equivalent File. The HILDA Survey features a complex, probabilistic sampling design (see for details Summerfield et al., 2015), and is largely representative of Australian households in 2001 in relation to labour market, housing, demographic and health variables (see Summerfield et al., 2015, pp. 117–118). Exceptions include individuals who are institutionalized and those who live in areas defined as "very remote" by the Australian Bureau of Statistics. All household members aged 15 or older who live in the selected households are asked to participate in the survey – in our analytical sample the age range is 15 to 101. In Wave 1 nearly 60% of in-scope households agreed to participate in the study, and interviews were collected with 92% of in-scope respondents in those households. All members of households in which at least one person provided an interview in Wave 1 of the survey were subsequently followed up over time. Any new household members are also interviewed and, if they marry or have a child with original sample members, they are also followed up over time as they move away into new households. Year-on-year respondent retention rates in the HILDA Survey are remarkably high for Australian and international standards, ranging between 87% and 97% (95% for the last study wave, Wave 14) (Summerfield et al., 2015). In all HILDA Survey waves, information is collected through a combination of face-to-face interviews and self-completion questionnaires (questions on mental disorders are contained within the former, whereas questions on summary mental health measures are contained within the latter).

The HILDA Survey is excellently suited to answer our research question because it features a unique combination of the following elements: (i) interviewer-reported paradata on IRRESI, (ii) multiple measures of respondent mental health and mental disorders, (iii) interviewer identifiers to account for unobserved interviewer effects, and (iv) repeated measurements from the same individuals over a long period of time to account for unobserved individual effects.

3.2 Dependent variables

Interviewer observations of respondent engagement with the survey interview. All interviewers in the HILDA Survey are professional interviewers from an external survey research company – The Nielsen Company up to Wave 9 (2009), and Roy Morgan Research thereafter – and are specifically trained to complete their HILDA Survey work. After the conclusion of each face-to-face interview, the interviewers are required to answer a set of questions about the interview situation. We peruse this information to derive three binary outcome variables tapping different IRRESI aspects.

The first variable uses information on interviewer answers to the question "Was the respondent suspicious about the study after the interview was completed?". The response "No, not at all suspicious" was recoded as 0, and the responses "Yes, somewhat suspicious" and "Yes, very suspicious" were recoded as 1. The second variable is derived using interviewers' answers to the question "In general, how would you describe the respondent's understanding of the questions?". The responses "excellent" and "very good" were recoded as 0, and the responses "fair", "poor" and "very poor" were recoded as 1. The third variable is based on interviewers' answers to the question "In general, how would you describe the respondent's co-operation during the interview?". The responses "excellent" and "very good" were recoded as 0, and the responses "fair", "poor" and "very poor" were recoded as 1. Hence, for the three binary outcome variables a value of 1 indicates a suboptimal interview outcome, and a value of 0 an optimal interview outcome.

The HILDA Survey question used to derive the first IRRESI outcome is based on a question included in the 1998 US Survey of Consumer Finances (Kennickell, Starr-McCluer, & Surette, 2000), whereas the HILDA Survey questions used to derive the second and third IRRESI outcomes were previously included in the British Household Panel Survey (Taylor, with Brice, Buck, & Prentice-Lane, 2010). Jointly, our three outcome variables provide complementary insights into overall IRRESI. In the HILDA Survey sample, in 2% of the person-year observations ($n = 3,964$) interviewers reported that respondents were suspicious of the study after the interview, in 4% ($n = 8,282$) that respondents had issues understanding the survey questions, and in 2% ($n = 3,210$) that respondents were uncooperative (Table 1).

Objective indicators of survey interview outcomes. In addition to interviewer-reported paradata, we also consider the associations between respondent mental health and mental disorders and three indicators of objective survey interview outcomes. The first indicator of OSIO is a dummy variable capturing panel attrition. This takes the value 1 if the respondent does not participate in the next wave of the HILDA Survey, and the value 0 otherwise. Note that this measure cannot be calculated for observations from the last survey wave used in our analyses (wave 14), as we cannot determine whether the respondent will or will not participate in the subsequent survey wave. On average, in about 7% of person-year observations ($n = 12,395$) HILDA Survey respondents do not participate in the following survey wave. The second OSIO measure is based on the degree of item-level missing data over the course of the face-to-face interview, where missing data items are defined as "don't know" answers, refusals, responses with missing information, or answers containing implausible values. To derive this variable, we first calculate the number of questions an individual was asked during the face-to-face components of the HILDA Survey interview – which is dependent on his/her individual circumstances – and the number of missing data items. We subsequently divide the number of data items by the total number of survey questions, and multiply the resulting figure by 100 to be expressed as a percentage. To be consistent across outcome variables, we then create a dummy variable taking the value 1 if 2% of more of the individual data items were missing, and the value 0 otherwise. We choose this arbitrary 2% threshold to isolate the tail of the distribution, so that in 10% of the person-year observations respondents score 1 in this variable ($n = 19,370$). Results were similar using other thresholds. Our third and final OSIO indicator is a dummy variable denoting whether or not respondents completed and returned their self-complete questionnaire (SCQ). As indicated in the HILDA Survey user manual, all individuals completing a face-to-face personal interview "are asked to complete a Self-Completion Questionnaire which the interviewer collects at a later date, or failing that, is returned by mail. This questionnaire comprises mainly attitudinal questions, many of which cover topics which respondents may feel slightly uncomfortable answering in a face-to-face interview" (Summerfield et al., 2015: 53). The value 1 is assigned to those person-year observations in which respondents failed to complete/return their SCQs (10%, $n = 20,897$), and the value 0 to those person-year observations in which respondents did complete/return their SCQs.

3.3 Explanatory variables

Mental health scales. As ours is an exploratory exercise, we use several measures of mental health and mental disorders available in the HILDA Survey. While there are differences and some potential overlap in what these measures capture, we expect that for all of them better mental health and the absence of mental disorders relates to better survey outcomes.

Our first mental health measure is the SF-36 Mental Health Inventory (MHI-5) (Ware & Sherbourne, 1992), which is available across all 14 waves of the HILDA Survey (2001–2014). The MHI-5 captures psychological well-being and the absence of psychological distress. It is constructed out of responses to 5 questions about how often in the past 4 weeks respondents had: "been a nervous person", "felt so down in the dumps that nothing could cheer them up", "felt calm and peaceful", "felt down" and "been a happy person". Possible responses are: "all of the time", "most of the time", "a good bit of the time", "some of the time", "a little of the time" and "none of the time". Following conventions in the literature, we rescaled the resulting MHI-5 index to range from 0 (worst mental health) to 100 (best mental health). In our HILDA Survey sample, the MHI-5 variable has a mean of 74.2, a standard deviation of 17.14, and its distribution covers the entire possible range of 0–100 (Table 1).

Our second mental health measure is the Kessler Psychological Distress Scale (K10). The K10 captures levels of non-specific psychological distress and depressive symptoms (Kessler et al., 2002), and is constructed out of responses to 10 questions about how often in the past 4 weeks respondents felt "tired for no good reason", "nervous", "so nervous that nothing could calm them down", "hopeless", "restless or fidgety", "so restless that they could not sit still", "depressed", "that everything was an effort", "so sad that nothing could cheer them up", and "worthless". Possible responses are: "all the time", "most of the time", "some of the time", "a little of the time" and "none of the time". When these are added up, the resulting K10 index ranges from 10 (lowest psychological distress) to 50 (highest psychological distress). Information on the K10 is available in HILDA Survey waves 7 (2007), 9 (2009), 11 (2011) and 13 (2013). In these data, the K10 has a mean of 15.72, a standard deviation of 6.3, and its distribution covers the entire possible range of 10–50 (Table 1).

Results using dichotomous versions of the MHI-5 and K10 based on critical thresholds (not shown but available upon request) are similar to those presented here. We retain the continuous-level summary mental health measures in the main models as they display more variance and are hence more informative.

Binary indicators of mental disorders. Using responses from a HILDA Survey multi-response question available in waves 3-14 (2003–2014), we construct three additional binary variables capturing more severe and long-lasting mental disorders. Specifically, HILDA Survey participants are asked whether they have "any long-term health condition, impairment or disability that restricts their everyday activities, and has lasted or is likely to last, for 6 months or more", while being shown a list of conditions in a showcard. The question wording and showcard were based on survey items included in the Australian Government Department of Family and Community Services General Customer Survey and the Australian Bureau of Statistics Survey of Training and Education. We consider three mental disorders: (i) "a mental illness that requires help or supervision", (ii) "difficulty learning or understanding things", and (iii) "a nervous or emotional condition that requires treatment". In the HILDA Survey data, respondents report having a mental illness requiring help/supervision in 2% of the person-year observations ($n = 2,601$), difficulty learning/understanding things in 1% of the person-year observations ($n = 2,226$), and a nervous/emotional condition requiring treatment in 4% ($n = 6,142$) of the person-year observations.

3.4 Analytic approach

We begin by estimating unadjusted logistic regression models without control variables on each of the three outcome variables. These unadjusted models give the "raw" associations between respondent mental health, mental disorders and survey outcomes, and take the form

$$\log\left(\frac{\Pr(O_{ijt}=1)}{1-\Pr(O_{ijt}=1)}\right) = \alpha + \beta_1 M_{ijt} + e_{ijt}, \quad , \qquad (1)$$

where the subscripts i, j and t refer to individual, interviewer, and time period, respectively; O is a dichotomous outcome variable capturing an aspect of IRRESI or OSIO; αa is the model's grand intercept; M is a given measure of respondents' mental health or mental disorders and βb_1its associated estimated coefficient; and e is the usual random error term in regression estimation. The results of these models are used to compute predicted probabilities for the outcome variables at different levels of the explanatory variables capturing mental health and disorders. These are helpful to determine the magnitude of the differences in IRRESI and OSIO across individuals with different mental health levels, or with and without mental disorders. Because the different measures of mental health and disorders (H) tap similar constructs and are sometimes highly correlated, we fit separate models for each of them.

We then estimate a second set of models to test whether the associations between mental health, mental disorders and survey outcomes are also apparent in the presence of confounders at the observation, individual and interviewer levels. If so, that would provide stronger evidence that the differences in IRRESI and OSIO are indeed due to respondents' mental health and disorders. However, we acknowledge that identifying causal relationships may not be possible with these observational data for reasons discussed below. Accounting for unobserved confounders is particularly important, as a degree of subjectivity is involved in interviewers' IRRESI reports, and because interviewers are pivotal to eliciting good quality survey outcomes (Durrant, Groves, Staetsky, & Steele, 2010). To accomplish this, we deploy three-level (multilevel) models, as these are the optimal way to model data in which person-year observations (Level 1) are nested within survey respondents (Level 2), who are in turn nested within survey interviewers (Level 3) (Lynn, Kaminska, & Goldstein, 2014; Vassallo, Durrant, Smith, & Goldstein, 2015). Further, the models allow for cross classification (i.e. non-pure nesting), given that the same interviewer can interview different respondents within and across survey waves, and that the same respondent can be interviewed by different interviewers over time (Browne, Goldstein, & Rasbash, 2001; Hill & Goldstein, 1998). Since the outcome variables are dichotomous, we estimate logistic regression models

$$\log\left(\frac{\Pr(O_{ijt}=1)}{1-\Pr(O_{ijt}=1)}\right) = \alpha + \beta_1 M_{ijt} + \gamma_1 X_{ijt} + \sum_{i}^{T} w_{ijt} u_{jt} + v_{ij} + e_{ijt} \quad . \qquad (2)$$

Here, the X_{ijt} is a vector of control variables capturing a comprehensive set of factors suspected to confound the associations of interest (see Table 1), and γ a transposed vector of their associated estimated coefficients; u_{jt} are the interviewer-level random effects capturing interviewer-specific unobserved heterogeneity; v_{ij} are the individual-level random effects capturing individual-specific unobserved heterogeneity; and e_{ijt} is the usual random error term in regression. The interviewer effect (u_{jt}) assigned to each respondent in this cross-classified model is a weighted average of the random effect for each of the interviewers with whom the respondent engaged over its participation in the panel, with weights (w_{ijt}) adding up to one (Durrant et al., 2010). The models were estimated using MLwiN 2.25 software and Markov Chain Monte Carlo (MCMC) methods (Rasbash & Browne, 2008). For ease of interpretation, we report the estimates of all logistic regression models as odds ratios. The data cannot be weighted for sample selection or attrition because of the complexity of its structure and the requisite modelling (particularly the cross-classification), and because the longitudinal weights in the HILDA Survey are only available for a balanced sample of respondents (which reduces the analytic sample by 80% and introduces further selectivity issues). As a robustness check, we re-estimated all models using the longitudinal weights as covariates, which bypasses the latter issue. The results were virtually identical to those presented here and are available from the authors upon request.

The next section presents our empirical findings.

4 Respondent mental health, mental disorders and survey outcomes

4.1 Unadjusted logistic regression models

Results from unadjusted logistic regression models are summarized in Table 2. Results in Column 1 to 3 are for models of IRRESI. Better mental health measured by the MHI-5 (OR = 0.995, $p < 0.001$) and the absence of psychological distress, measured by the K10 (OR = 1.014, $p < 0.001$) reduce the likelihood of interviewers reporting that respondents were suspicious of the study after the interview (Column 1). However, none of the three mental disorders is statistically significantly related to this outcome. Lower scores in the MHI-5 (OR = 0.981, $p < 0.001$) and higher K10 scores (OR = 1.059, $p < 0.001$) are associated with a higher likelihood of interviewers rating respondents as experiencing issues understanding the survey questions, and so are the presence of a mental illness requiring help/supervision (OR = 3.577, $p < 0.001$), difficulty learning/understanding things (OR = 11.339, $p < 0.001$), and having a nervous/emotional condition (OR = 2.115, $p < 0.001$) (Column 2). Lower scores in the MHI-5 (OR = 0.988, $p < 0.001$), higher K10 scores (OR = 1.025, $p < 0.001$),

Table 2
Unadjusted logistic regression models, odds ratios

	Interviewer observations			Survey outcomes		
	Suspicious of interview (1)	Poor question understanding (2)	Lack of cooperation (3)	Panel attrition (4)	Item-level missing data (5)	No SCQ (6)
Mental health scales						
SF-36 Mental Health Inventory[a]	0.995***	0.981***	0.988***	0.990***	0.991***	-[d]
N(obs)=177,973 / N(ind)=27,165 / N(int)=632						
K10 Psychological Distress Scale[b]	1.014*	1.059***	1.025***	1.031***	1.027***	-[d]
N(obs)=53,145 / N(ind)=20,140 / N(int)=360						
Mental disorders						
Mental illness that requires help/supervision[c]	1.082	3.577***	2.264***	1.321***	2.005***	1.444***
N(obs)=172,962 / N(ind)=26,445 / N(int)=556						
Difficulty learning/understanding things[c]	1.100	11.339***	3.139***	1.650***	3.284***	2.656***
N(obs)=172,962 / N(ind)=26,445 / N(int)=556						
Nervous/emotional condition requiring treatment[c]	1.025	2.115***	1.320**	1.057	1.661***	1.202***
N(obs)=172,962 / N(ind)=26,445 / N(int)=556						

HILDA Survey data, Australia. [a] Data for years 2001–2014. [b] Data for years 2007, 2009, 2011 & 2013. [c] Data for years 2003–2014.
[d] Not applicable, as K10 and MHI-5 data are contained within the SCQ.
* $p < .05$ ** $p < .01$ *** $p < .001$

and having a mental illness requiring help/supervision (OR = 2.264, $p < 0.001$), difficulty learning/understanding things (OR = 3.319, $p < 0.001$) and a nervous/emotional condition (OR = 1.320, $p < 0.01$) are also related to increased odds of interviewers reporting uncooperativeness amongst survey respondents (Column 3).

Results for Columns 4 to 6 are for models of OSIO. Lower scores in the MHI-5 (OR = 0.990, $p < 0.001$), higher K10 scores (OR = 1.031, $p < 0.001$), having a mental illness requiring help/supervision (OR = 1.321, $p < 0.001$), and having difficulty learning/understanding things (OR = 1.650, $p < 0.001$) are all related to the probability of panel attrition (Column 4). However, having a nervous/emotional condition (OR = 1.057, $p < 0.1$) is not. Lower scores in the MHI-5 (OR = 0.991, $p < 0.001$) and higher K10 scores (OR = 1.027, $p < 0.001$) are associated with the likelihood of missing survey items (Column 5). Similarly, having a mental illness requiring help/supervision (OR = 2.005, $p < 0.001$), having difficulty learning/understanding things (OR = 3.284, $p < 0.001$), and having a nervous/emotional condition (OR = 1.661, $p < 0.001$) significantly increase such likelihood. Finally, mental illnesses requiring help/supervision (OR = 1.444, $p < 0.001$), difficulty learning/understanding things (OR = 2.656, $p < 0.001$), and nervous/emotional conditions (OR = 1.202, $p < 0.001$) are all associated with the likelihood that respondents do not complete/return their SCQs (Column 6).

Altogether, 25 of the 27 coefficients of interest on the mental health and mental disorder measures in these unadjusted logit models are statistically significant. All of them follow the hypothesized direction, indicating that poor mental health and the presence of mental disorders is associated with poorer IRRESI and OSIO.

4.2 Predicted probabilities

To get a sense of the magnitude of the estimated effects in the unadjusted logistic regression models, Table 3 presents the predicted probabilities at the 10th, 25th, 50th, 75th and 90th percentiles of the continuous mental health measures (the MHI-5 and K10), and at the values 0 and 1 of the binary mental disorder measures.

The magnitude of association between the mental health and mental disorder variables and the outcome variable capturing respondents being suspicious of the study is very small. Similarly, the magnitude of association between the summary mental health variables and the outcome variable capturing interviewer perceptions of lack of cooperation by respondents is also small. However, such magnitude is bigger for the mental disorder variables: while 1.5% of people with no mental disorders are predicted to be deemed uncooperative by interviewers, the rates are two-to-three times greater amongst people with a mental illness requiring help (3.3%) and with learning/understanding difficulties (4.5%). Amongst the IRRESI measures, effect sizes are greatest for the outcome variable capturing interviewer reports of poor question comprehension amongst respondents. For example, 4.6% of individuals in the 10th percentile of the MHI-5 distribution are predicted to be rated by interviewers as being suspicious of the study, compared to 2.1% of individuals in the 90th percentile of the MHI-5 distribution. Amongst the mental disorders, results for this outcome are striking: 3.6% to 3.8% of respondents without mental disorders are predicted to be reported by interviewers as having trouble understanding the survey questions, compared to 7.8% of respondents with nervous/emotional problems, 12.5% of respondents with a mental illness requiring help, and 30% of respondents with learning/understanding difficulties.

Turning now our attention to the measures of OSIO, we observe that the magnitude of association between the mental health and disorder variables and the outcome variable capturing panel attrition is moderate. As an illustration, 6.9% of individuals in the 10th percentile of the MHI-5 distribution are predicted to stop responding to the panel in the next wave, compared to 4.8% of individuals in the 90th percentile of the MHI-5 distribution. Effect sizes for two of the mental disorders are also non-negligible: while 6.2% of people with no mental disorders attrite from the study, 8.1% of people with mental illnesses requiring help/supervision and 9.8% of people with difficulty learning/understanding things do so. The magnitude of association between the summary mental health variables and the outcome variable capturing the rate of item-level missing data is moderate, but such magnitude is again bigger for mental disorders: while 10.8–10.9% of people with no mental disorders are predicted to score one in such variable, the rates are higher amongst people with nervous/emotional conditions (12.9%), mental illnesses requiring help/supervision (15.1%) and learning/understanding difficulties (24.4%). Finally, there are also elevated rates of missing SCQs amongst individuals with mental disorders. Such rates are roughly 10.2% amongst people with no such disorders, but 15.8% amongst those with nervous/emotional conditions, 18.6% for those with mental illnesses requiring help/supervision, and a striking 27% for people with learning/understanding difficulties.

4.3 Three-level, cross-classified logistic regression models

Results from our three-level, cross-classified models are summarized in Table 4. These are revealing as to whether or not the measures of mental health and mental disorders are associated with IRRESI and OSIO when adjusting for observed and unobserved observation- and individual-level factors, as well as unobserved interviewer-level effects. We note, however, that direct comparisons of odds ratios between adjusted and unadjusted logit models are inappropriate due to the “scaling problem” (Mood, 2010).

Table 3
Predicted probabilities from unadjusted logistic regression models (in %)

	Percentile					Disorder	
	10th	25th	50th	75th	90th	0	1
Interviewer assessment: Respondent was suspicious of interview							
SF-36 Mental Health Inventory[a]	1.9	1.8	1.7	1.6	1.6		
K10 Psychological Distress Scale[c]	1.1	1.1	1.2	1.2	1.4		
Mental illness requiring help[c]						1.6	1.8
Difficulty learning/understanding[c]						1.6	1.8
Nervous/emotional condition[c]						1.6	1.7
Interviewer assessment: Respondent displayed poor question understanding							
SF-36 Mental Health Inventory[a]	4.6	3.6	2.7	2.3	2.1		
K10 Psychological Distress Scale[c]	2.0	2.1	2.5	3.1	4.4		
Mental illness requiring help[c]						3.8	12.5
Difficulty learning/understanding[c]						3.8	30.0
Nervous/emotional condition[c]						3.6	7.8
Interviewer assessment: Respondent was uncooperative							
SF-36 Mental Health Inventory[a]	1.6	1.4	1.2	1.0	1.0		
K10 Psychological Distress Scale[b]	0.9	0.9	1.0	1.1	1.2		
Mental illness requiring help[c]						1.5	3.3
Difficulty learning/understanding[c]						1.5	4.5
Nervous/emotional condition[c]						1.5	2.0
Survey outcome: Panel attrition							
SF-36 Mental Health Inventory[a]	6.9	6.2	5.4	5.0	4.8		
K10 Psychological Distress Scale[c]	4.1	4.2	4.6	5.2	6.2		
Mental illness requiring help[c]						6.2	8.1
Difficulty learning/understanding[c]						6.2	9.8
Nervous/emotional condition[c]						6.2	6.6
Survey outcome: Item-level missing data							
SF-36 Mental Health Inventory[a]	10.2	9.3	8.2	7.6	7.4		
K10 Psychological Distress Scale[c]	7.7	7.9	8.5	9.3	10.8		
Mental illness requiring help[c]						10.9	15.1
Difficulty learning/understanding[c]						10.8	24.4
Nervous/emotional condition[c]						10.9	12.9
Survey outcome: No SCQ							
SF-36 Mental Health Inventory[a]	-[d]	-[d]	-[d]	-[d]	-[d]		
K10 Psychological Distress Scale[c]	-[d]	-[d]	-[d]	-[d]	-[d]		
Mental illness requiring help[c]						10.2	18.6
Difficulty learning/understanding[c]						10.1	27.0
Nervous/emotional condition[c]						10.2	15.8

HILDA Survey data, Australia. [a] Data for years 2001–2014. [b] Data for years 2007, 2009, 2011 & 2013.
[c] Data for years 2003–2014. [d] Not applicable, as K10 and MHI-5 data are contained within the SCQ.
$^{*}\ p < .05$ $^{**}\ p < .01$ $^{***}\ p < .001$

Table 4
Cross-classified multilevel logistic regression models, odds ratios

	Interviewer observations			Survey outcomes		
	Suspicious of interview (1)	Poor question understanding (2)	Lack of cooperation (3)	Panel attrition (4)	Item-level missing data (5)	No SCQ (6)
Mental health scales						
SF-36 Mental Health Inventory[a]	0.995***	0.986***	0.988***	0.989***	0.992***	-[d]
N(obs)=177,973 / N(ind)=27,165 / N(int)=632						
K10 Psychological Distress Scale[b]	1.011	1.054***	1.018*	1.026***	1.040***	-[d]
N(obs)=53,145 / N(ind)=20,140 / N(int)=360						
Mental disorders						
Mental illness that requires help/supervision[c]	1.052	2.999***	2.018***	1.472***	2.371***	1.407***
N(obs)=172,962 / N(ind)=26,445 / N(int)=556						
Difficulty learning/understanding things[c]	1.161	7.194***	2.358***	1.464***	4.430***	2.050***
N(obs)=172,962 / N(ind)=26,445 / N(int)=556						
Nervous/emotional condition requiring treatment[c]	1.029	1.680***	1.284*	1.250***	2.606***	1.422***
N(obs)=172,962 / N(ind)=26,445 / N(int)=556						

HILDA Survey data, Australia; Models control for all variables shown in Table 1; Full tables of coefficients are available from the authors upon request.
[a] Data for years 2001–2014. [b] Data for years 2007, 2009, 2011 & 2013. [c] Data for years 2003–2014. [d] Not applicable, as K10 and MHI-5 data are contained within the SCQ.
* $p < .05$ ** $p < .01$ *** $p < .001$

Altogether, 24 of the 25 coefficients on the mental health and mental disorder variables found to be statistically significant in the unadjusted models are also statistically significant in the adjusted models. As an exception, the K10 is no longer statistically significantly related to interviewer reports of respondents being suspicious of the study (Column 1). 14 mental of these estimates have a smaller magnitude than the respective estimates in the unadjusted models, 10 have a bigger magnitude, and 2 are of the same magnitude. All 24 statistically significant estimated effects in the adjusted models go in the hypothesized direction, indicating that poor mental health and the presence of mental disorders are associated with poorer IRRESI and worse OSIO.

5 Discussion and conclusion

5.1 Summary of study aims and findings

Despite increasing Government expenditure in health services, the prevalence of poor mental health and mental disorders in countries such as Australia has remained stable or even increased (Department of Health and Ageing., 2013). In this context, gaining a robust understanding of the predictors and consequences of poor mental health and the emergence of mental disorders is a fundamental goal of contemporary health research, and findings from survey research are frequently used to inform preventive and remedial health policy and practice. Yet, there is virtually no empirical evidence comparing how individuals with poorer and better mental health, or with and without mental disorders, engage with survey interviews. In this paper, we contributed to filling this gap in knowledge using a unique panel dataset that is largely representative of the Australian population and state-of-the-art multilevel regression models.

Drawing on information processing theory, we hypothesized that individuals with poor mental health and with mental disorders would display poorer outcomes concerning (i) interviewer reports of respondent engagement with the survey interview (IRRESI), and (ii) objective survey interview outcomes (OSIO). Our empirical findings are consistent with this expectation: individuals with poor mental health or who experience mental disorders are more likely to receive low IRRESI and experience worse OSIO. These associations were visible across a range of IRRESI and OSIO outcomes (interviewers reports that respondents were suspicious of the study, had issues understanding survey questions and were uncooperative, panel attrition, item-level missing data and failure to return/complete the SCQ) and health measures (MHI-5, K10, and three indicators of mental disorders). However, the magnitude of the associations varied across models. Differences in IRRESI and OSIO by mental health and mental disorders were more pronounced and more often statistically significant for the outcome variable measuring interviewer ratings of question comprehension, and the three indicators of OSIO. They were also visibly larger for the variables capturing mental disorders than the summary mental health measures. Statistically significant associations between the measures of mental health and mental disorders and the IRRESI and OSIO outcome variables were also apparent in multivariate logistic regression models accounting for observed and unobserved observation- and individual-level factors, as well as unobserved interviewer-level effects. This suggests that such associations are not the product of confounders.

5.2 Implications for survey practice

The observed deficits in IRRESI and OSIO amongst respondents with poor mental health and who experience mental disorders constitute new knowledge, and add to existing evidence indicating that ill mental health is a precursor of non-participation in surveys and attrition from prospective surveys (Australian Bureau of Statistics., 2009; Watson & Wooden, 2009). The results for the IRRESI measures suggest that, if professionally-trained interviewers are unbiased in their assessments (see discussion below), the validity and reliability of the resulting survey data may be comparatively lower amongst respondents with poor mental health or with mental disorders – as suspiciousness, uncooperativeness and poor question understanding are often deemed as significant barriers to the provision of accurate survey responses (Groves, 2004; Groves et al., 2009; Schwarz, 2007). However, a degree of caution needs to be exerted before concluding this, as there is little research on whether or not interviewer reports of respondents' behaviors during the interview are good predictors of data quality (Plewis, Calderwood, & Mostafa, 2017). In fact, some research has found substantial measurement error in interviewer reports of other domains, such as respondents' ethnicity, neighborhood appearance or dwelling type (Casas-Cordero, Kreuter, Wang, & Babey, 2013; Sinibaldi, Durrant, & Kreuter, 2013).

Our second set of results for the OSIO measures follows a similar pattern. Since these are objective indicators of survey engagement rather than subjective reports from interviewers, this lends additional support to our first set of results. The OSIO results are also of relevance to survey data users and collectors in their own right. The findings for panel attrition suggest that the samples of longitudinal surveys may become progressively skewed towards individuals with high levels of mental health and no mental disorders. While this is an issue that could be corrected (or at least ameliorated) by incorporating mental health and mental disorder information in the derivation of longitudinal survey weights, doing so is not customary. Our OSIO results also suggest that complete-case survey analyses as well as analyses involving variables contained within SCQs are likely to be based on samples which are biased towards respondents with good mental health and no mental disorders. The latter highlights the importance of empirically grappling with the issue of mental health and mental disorders predicting data availability (e.g., via multiple imputation, weighting or sample selection models).

It is nevertheless important to emphasize that the rates of mental disorders and some of the IRRESI and OSIO variables in the HILDA Survey are very low. Hence, the pattern of results reported here is unlikely to pose serious issues within the context of this specific survey, particularly to the "average" data user. However, there may be more important consequences for other data collections; particularly surveys explicitly aimed at gathering information on individuals with mental health issues (e.g. medical expenditure surveys), or surveys focused on population subgroups in which mental disorders are prevalent (e.g. elderly people, crime victims, war veterans, or sexual minorities). This poses questions about what could be done to improve the ways in which individuals with poor mental health and with mental disorders engage with the survey interview process. One possibility is that, if information on mental health and/or mental disorders is screened, collected early on in the study, or in a previous wave of a longitudinal survey, such information could be used to tailor the survey experience of such individuals. Survey instruments and protocols could be selectively adapted in ways that mitigate the barriers leading to poor survey outcomes amongst this subpopulation, but this requires further research that can ascertain the nature of such barriers. This could take the form of studies involving cognitive interviewing techniques or detailed examinations of interviewer-interviewee interactions (Hartley & MacLean, 2006). In addition, survey practitioners could provide some basic training to survey interviewers on how to approach and interact with respondents with poor mental health or who experience mental disorders (Becker, Roberts, Morrison, & Silver, 2004) – similar to the cultural-competence training provided to people who frequently work with vulnerable population groups, such as ethnic minorities and LGBT people (Betancourt, Green, Carrillo, & Ananeh-Firempong, 2003; Westerman, 2004).

5.3 Study limitations and avenues for further research

Despite the uniqueness and relevance of our findings, our study suffers from several data-driven shortcomings which point towards avenues for future methodological refinement. First, individuals with poor mental health and mental disorders are less likely to participate in surveys and the HILDA Survey sample does not cover the institutionalized population (e.g. people living in elderly homes, prisons or mental facilities), which are likely to suffer from more and more intense mental health problems. As a result, it is likely that individuals with poor mental health and mental disorders in our sample are "positively selected". If so, the negative effects of mental health and mental disorders on IRRESI and OSIO that we report may be conservative (i.e. downward-biased) estimates of the true relationships. Second, while our research leverages unique data from the HILDA Survey and the available summary measures of mental health are the gold standard in survey research, the measures of mental disorders do not correspond to those used in other widespread survey instruments designed to measure self-reported diagnostic disorders, such as the Composite International Diagnostic Interview (CIDI). They are also very coarse, failing to reflect the complexity of mental disorders reflected in the DSM-5 of the International Classification of Diseases (ICD-10). As a result, the broad results that we present here may mask substantial heterogeneity and may differ when other measurement tools for mental disorders are employed. Further research using alternative measures of mental disorders is thus warranted. Third, we do not claim that the associations we find are causal. Particularly, some of the estimated effect of respondent mental health and mental disorders on IRRESI may be due to reverse causation. That is, we cannot rule out that interviewers' attitudes towards mental health (e.g. the degree to which they stigmatize individuals with poor mental health) color their assessments of respondent engagement with the survey when they encounter respondents with ill mental health. For example, some interviewers may feel uncomfortable interacting with respondents who display cues of having poor mental health or mental disorders, and give artificially poor assessments due to their own prejudice. In fact, interviewers may be aware of the respondents' mental health and mental disorders through their knowledge of respondents' survey answers. While our model incorporates unobserved interviewer effects to minimize the potential bias, this may be insufficient to fully account for it. Improving our research in this direction would probably entail the collection of new fit-for-purpose data, e.g. experimental data manipulating interviewer perceptions of the mental health and mental disorders of survey respondents. However, the fact that similar associations were found for objective survey outcomes lends credibility to our IRRESI results.

5.4 Concluding remarks

While surveys are powerful means by which to gather evidence to inform the development of mental health policies, data collectors, researchers and policymakers need to remain cognizant of potential issues emerging from differences in the ways in which individuals with poorer and better mental health, and with and without mental disorders, engage in social surveys. Our findings paint a somewhat positive picture, suggesting that there are only small differences in such processes. Yet, more research on whether or not, and if so how, poor mental health and mental disorders are related to survey outcomes and survey data quality is needed.

6 Acknowledgments

The authors would like to thank Walter Forrest and Stefanie Plage for helpful comments and suggestions on an earlier draft. This research was supported by the Australian Research Council Centre of Excellence for Children and Families over the Life Course (project number CE140100027). The views expressed herein are those of the authors and are not necessarily those of the Australian Research Council. This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the author and should not be attributed to either DSS or the Melbourne Institute.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington: American Psychiatric Publishing.

Australian Bureau of Statistics. (2009). *4327.0 – National Survey of Mental Health and Wellbeing: users' guide, 2007.* Canberra: Australian Bureau of Statistics.

Becker, H., Roberts, G., Morrison, J., & Silver, J. (2004). Recruiting people with disabilities as research participants: challenges and strategies to address them. *Mental Retardation*, *42*(6), 471–475.

Betancourt, J. R., Green, A. R., Carrillo, J. E., & Ananeh-Firempong, O. (2003). Defining cultural competence: a practical framework for addressing racial/ethnic disparities in health and health care. *Public Health Reports*, *118*(4), 293–302.

Borgers, N., de Leeuw, E., & Hox, J. (2000). Children as respondents in survey research: cognitive development and response quality 1. *66*(1), 60–75.

Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling*, *1*(2), 103–124.

Casas-Cordero, C., Kreuter, F., Wang, Y., & Babey, S. (2013). Assessing the measurement error properties of interviewer observations of neighbourhood characteristics. *Journal of the Royal Statistical Society: Series A*, *176*(1), 227–249.

Department of Health and Ageing. (2013). *National mental health report 2013: tracking progress of mental health reform in Australia 1993–2011.* Canberra: Commonwealth of Australia.

Durrant, G. B., Groves, R. M., Staetsky, L., & Steele, F. (2010). Effects of interviewer attitudes and behaviours on refusal in household surveys. *Public Opinion Quarterly*, *74*(1), 1–36. doi:doi:10.1093/poq/nfp098

Eren, N. & Şahin, S. (2016). An evaluation of the difficulties and attitudes mental health professionals experience with people with personality disorders. *Journal of Psychiatric and Mental Health Nursing*, *23*(1), 22–36. doi:10.1111/jpm.12257

Esposito, J. L. & Jobe, J. B. (1991). A general model of the survey interaction process. In *In proceedings of the bureau of the census seventh annual research conference proceedings, march 19* (pp. 537–560). Washington, DC: US Bureau of the Census.

Groves, R. M. (2004). *Survey errors and survey costs.* Hoboken: John Wiley & Sons, Inc.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, N.J.: Wiley.

Hartley, S. L. & MacLean, W. E. (2006). A review of the reliability and validity of Likert-type scales for people with intellectual disability. *Journal of Intellectual Disability Research*, *50*(11), 813–827. doi:10.1111/j.1365-2788.2006.00844.x

Hill, P. W. & Goldstein, H. (1998). Multilevel modeling of educational data with cross-classification and missing identification for units. *Journal of Educational and Behavioral Statistics*, *23*(2), 117–128.

Kennickell, A. B., Starr-McCluer, M., & Surette, B. J. (2000). Recent changes in US family finances: results from the 1998 Survey of Consumer Finances. *Federal Research Bulletin*, *86*, 1–29.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., … Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *32*(6), 959–976.

Koenen, K. C., Moffitt, T. E., Roberts, A. L., Martin, L. T., Kubzansky, L., Harrington, H., … Caspi, A. (2009). Childhood IQ and adult mental disorders: a test of the cognitive reserve hypothesis. *American Journal of Psychiatry*, *166*(1), 50–7. doi:10.1176/appi.ajp.2008.08030343

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236. doi:10.1002/acp.2350050305

Kruijshaar, E. M., Barendregt, J., Vos, T., de Graaf, R., Spijker, J., & Andrews, G. (2005). Lifetime prevalence estimates of major depression: an indirect estimation method and a quantification of recall bias. *European Journal of Epidemiology*, *20*(1), 103–111.

Link, B. G., Phelan, J. C., Bresnahan, M., Stueve, A., & Pescosolido, B. A. (1999). Public conceptions of mental illness: labels, causes, dangerousness, and social distance. *American Journal of Public Health*, *89*(9), 1328–1333.

Lynn, P., Kaminska, O., & Goldstein, H. (2014). Panel attrition: how important is interviewer continuity? *Journal of Official Statistics*, *30*(3), 443–457. doi:10.2478/jos-2014-0028

Meisenberg, G. & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, *44*(7), 1539–1550. doi:10.1016/j.paid.2008.01.010

Mood, C. (2010). Logistic regression: why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, *26*(1), 67–82. doi:10.1093/esr/jcp006

Patten, S. B. (2003). Recall bias and major depression lifetime prevalence. *Social Psychiatry and Psychiatric Epidemiology*, *38*(6), 290–6. doi:10.1007/s00127-003-0649-9

Perales, F., Baffour, B., & Mitrou, F. (2015). Ethnic differences in the quality of the interview process and implications for survey analysis: the case of indigenous Australians. *10*(6), e0130994. doi:10.1371/journal.pone.0130994

Plewis, I., Calderwood, L., & Mostafa, T. (2017). Can interviewer observations of the interview predict future response? *Methods, Data, Analyses: A Journal for Quantitative Methods and Survey Methodology*, *11*(1).

Rasbash, J. & Browne, W. J. (2008). Non-hierarchical multilevel models. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 301–334). New York: Springer.

Satcher, D. (2000). Mental health: a report of the surgeon general – executive summary. *Professional Psychology: Research and Practice*, *31*(1), 5–13.

Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, *21*(2), 277–287. doi:10.1002/acp.1340

Sigelman, C. K., Schoenrock, C. J., Spanhel, C. L., Hromas, S. G., Winer, J. L., Budd, E. C., & Martin, P. W. (1980). Surveying mentally retarded persons: responsiveness and response validity in three samples. *American Journal of Mental Deficiency*, *84*(5), 479–86.

Sinibaldi, J., Durrant, G. B., & Kreuter, F. (2013). Evaluating the measurement error of interviewer observed paradata. *Public Opinion Quarterly*, *77*(S1), 173–193.

Summerfield, M., Freidin, S., Hahn, M., Li, N., Macalalad, N., Mundy, L., … Wooden, M. (2015). *HILDA user manual – release 14.* Melbourne: Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Taylor, M. F., with Brice, J., Buck, N., & Prentice-Lane, E. (Eds.). (2010). *British Household Panel Survey user manual volume A: introduction, technical report and appendices.* Colchester: University of Essex.

The Melbourne Institute of Applied Economic and Social Research (MIAESR). (2014). Household, Income and Labour Dynamics in Australia (HILDA) survey, 2001-2014. Retrieved from http://melbourneinstitute.unimelb.edu.au/hilda.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* Cambridge: Cambridge University Press.

Treloar, A. J. C. (2009). A qualitative investigation of the clinician experience of working with borderline personality disorder. *New Zealand Journal of Psychology*, *38*(2), 30–34.

Vassallo, R., Durrant, G. B., Smith, P. W. F., & Goldstein, H. (2015). Interviewer effects on non-response propensity in longitudinal surveys: a multilevel modelling approach. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *178*(1), 83–99. doi:10.1111/rssa.12049

Ware, J. E. & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36): I. conceptual framework and item selection. *Medical Care*, *30*(6), 473–483.

Watson, N. & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 157–181). Chichester: John Wiley and Sons.

Watson, N. & Wooden, M. (2012). The HILDA survey: a case study in the design and development of a successful household panel survey. *Longitudinal and Life Course Studies*, *3*(3), 369–381. doi:http://dx.doi.org/10.14301/llcs.v3i3.208

Westerman, T. (2004). Engagement of indigenous clients in mental health services: what role do cultural differences play? *Australian e-Journal for the Advancement of Mental Health*, *3*(3), 88–93.

World Health Organization. (2004). *Promoting mental health: concepts, emerging evidence, practice: summary report.* Geneva: WHO.

World Health Organization. (2008). *The global burden of disease: 2004 update.* Geneva: WHO.

World Health Organization. (2010). *Mental health and development: targeting people with mental health conditions as a vulnerable group.* Geneva: WHO.

World Health Organization. (2017). Mental health of older adults. fact sheet. Retrieved from http://www.who.int/mediacentre/factsheets/fs381/en/