lifecoursecentre

ARC Centre of Excellence for Children and Families over the Life Course

Life Course Centre Working Paper Series

Respondent Mental Wellbeing and Interviewer Ratings of the Quality of the Survey Interview

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

Bernard Baffour
Institute for Social Science Research, The University of Queensland

No. 2017-02
January 2017

The University of Queensland Australia
The University of Melbourne
The University of Western Australia
The University of Sydney

Australian Government
Australian Research Council

NON-TECHNICAL SUMMARY

Mental health conditions are amongst the largest causes of disease burden at a global level, and understanding the predictors and consequences of ill mental health is a fundamental goal of health research, policy and practice. Many studies of mental health rely on the analysis of population surveys. However, this research makes one important assumption, namely that the accuracy of the information gathered in surveys is comparable for individuals with low and high levels of mental health. This is problematic, as there are reasons to expect poorer survey interview outcomes amongst individuals with ill mental health, which may in turn lead to less accurate responses to survey questions.

In this study, we fill a gap in knowledge by comparing interviewer ratings of the quality of the survey interview (IRQSI) between respondents with poorer and better mental health. We consider three aspects of IRQSI: (i) interviewer ratings of survey respondents being suspicious of the study, (ii) having issues understanding the survey questions, and (iii) being uncooperative. Survey methodology manuals emphasize the importance of respondent trust, cooperation and understanding in the survey interview situation, as poor performance in these dimensions may affect survey estimates by leading to higher missing data, measurement error and report bias.

Our findings are consistent with expectations: individuals with poorer mental health are more likely to display low IRQSI. These associations were visible across a range of IRQSI outcomes and measures of mental health and disorders. These observed deficits in IRQSI amongst respondents with poor mental health constitute new and important knowledge, with implications for how researchers undertake survey research on mental health and how they interpret the results. To the extent that professionally-trained interviewers are accurate in their assessments, this finding is suggestive that the accuracy of the resulting survey data is comparatively lower amongst respondents with poor mental health. Hence, it is possible that survey analyses of individuals with poor mental health produce unreliable results, which poses a challenge to the usefulness of findings generated using survey data to inform the design of evidence-based mental health policy.

We conclude that, while surveys are powerful means by which to gather evidence to inform the development of health policies, it is not clear that researchers and policymakers should take the accuracy of survey data generated from respondents with ill mental health for granted. More research aimed at comparing how individuals with poorer and better mental health engage in the survey process, and whether and how their poor mental health is related to the quality of the information retrieved from these individuals is sorely needed.

ABOUT THE AUTHORS

Francisco Perales is Senior Research Fellow in Family Dynamics in the Institute for Social Science Research at The University of Queensland. His recent work has been on social disadvantage, gender inequalities, life course transitions, subjective wellbeing, and the use of household panel surveys within sociology. His recent research has been published in outlets such as *Journal of Marriage and Family*, *Social Forces*, *Social Science Research*, and *European Sociological Review*. Email: f.perales@uq.edu.au

Bernard Baffour is Research Fellow in Social Statistics at the Institute for Social Science Research at The University of Queensland. His primary research interest has been on census and survey methodology, including the areas of non-response, weighting, survey quality and dual frame sampling, and recently published an article in the *Journal of Official Statistics* on the optimal weighting approaches for dual frame sampling in Australia. Other publications include articles in *Journal of the Royal Statistical Society* and *International Statistical Review*. Email: b.baffour@uq.edu.au

ACKNOWLEDGEMENTS: The authors would like to thank Ella Kuskoff and Nicole Kapelle for research assistance, and 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.

lifecoursecentre

(ARC Centre of Excellence for Children and Families over the Life Course)
Institute for Social Science Research, The University of Queensland (administration node)
UQ Long Pocket Precinct, Indooroopilly, Qld 4068, Telephone: +61 7 334 67477
Email: lcc@uq.edu.au, Web: www.lifecoursecentre.org.au

Abstract

Mental health conditions 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 often rely on surveys. However, there is scarce evidence about whether or not the accuracy of information gathered in face-to-face surveys differs for respondents with good and poor mental health. We examine the associations between participant mental health and interviewer ratings of the quality of the survey interview using 14 years (2001-2014) of annual, nationally-representative, Australian panel data (n~200,000). We find that individuals with poorer mental health are generally more likely than individuals with better mental health to be deemed by interviewers as being suspicious of the study, experiencing issues understanding survey questions, and being uncooperative. These associations are apparent in models that control for observable and unobservable observation- and individual-level factors, as well as unobserved interviewer-level effects. These findings suggest that survey data collected from individuals with poor mental health may be comparatively inaccurate, which has implications for how researchers undertake and interpret the results of survey research on mental health.

Keywords: mental health; mental conditions; interviewer observations; survey data quality; multilevel models; panel data; Australia

1 Introduction

Mental health conditions are amongst the largest causes of disease burden at a global level (World Health Organization 2004, 2008). As a result, understanding the predictors and consequences of ill mental health is a fundamental goal of health research, policy and practice, and has been the focus of a wealth of interdisciplinary research (Power 2010, Aneshensel, Phelan, and Bierman 2013). Many contemporary studies of mental health rely on the analysis of population surveys. For example, a Scopus search for research articles published in 2015 in which the terms "mental health" and "survey" appear in the article title, abstract or keywords yields 2,196 items. Recent studies have used survey data to evaluate how factors as diverse as marital loss (Hewitt, Turrell, and Giskes 2012), physical activity (Perales, del Pozo-Cruz, and del Pozo-Cruz 2014), financial strain (Dijkstra-Kersten et al. 2015), workplace bullying (Lahelma et al. 2012), and housing and neighborhood quality (Jones-Rounds, Evans, and Braubach 2014) affect individuals' mental health, and to ascertain how individuals' mental health is in turn associated with consequences across diverse life domains, including educational attainment (Johnston et al. 2014), labor market outcomes (Rudolph and Eaton 2016) and parenting practices (Tzoumakis, Lussier, and Corrado 2015).

Inadvertently, this body of research relies on one important assumption, namely that the accuracy of the information gathered in surveys is comparable for individuals with low and high levels of mental health. This applies to the accuracy of survey responses to routinely-collected background questions (e.g. questions on education, employment, income) and of responses to more specific survey modules. In practice, there is no empirical evidence about whether or not this assumption holds. This is problematic, as there are reasons to expect poorer survey interview outcomes amongst individuals with ill mental health, which may in turn lead to less accurate responses to survey questions.

In this study, we fill a gap in knowledge by comparing interviewer ratings of the quality of the survey interview (IRQSI) between respondents with poorer and better mental health. We consider three aspects of IRQSI: (i) interviewer ratings of survey respondents being suspicious of the study, (ii) having issues understanding the survey questions, and (iii) being uncooperative. Survey methodology manuals emphasize the importance of respondent trust, cooperation and understanding in the survey interview situation (Groves 2004, Groves et al. 2009). Poor performance in these dimensions may affect survey estimates by leading to higher item missing data, measurement error and report bias. For instance, suspicious respondents will be more likely to refuse to answer or lie about certain survey questions, particularly those perceived to be sensitive; uncooperative respondents may exert 'satisficing' in multi-response questions or intentionally provide false responses to reduce the interview length; and respondents experiencing comprehension issues may inadvertently provide inaccurate, incomplete or erroneous answers. In panel studies, low IRQSI has been shown to predict loss to follow-up (Watson and Wooden 2009).

While some studies from assorted fields of inquiry have touched upon the associations between mental wellbeing, participation in survey research and response outcomes, none has done so from the prism of interviewer observations. To fill this gap in knowledge, we assess differences in IRQSI by individuals' mental health using 14 years of annual, nationally representative, panel data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. We find robust evidence of associations between IRQSI and survey measures of individual mental health and disorders, which has implications for the collection, analysis and interpretation of survey data from individuals with poor mental health.

2 Background

Information processing theory perspectives in survey methodology argue 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 and Rasinski 2000, p.8). It has been argued that socio-demographic factors, such as age or cultural background, can affect how survey respondents engage in each of these phases by influencing individuals' capabilities and schemata (Groves 2004, Groves et al. 2009, Tourangeau et al. 2000). Similar arguments have been made about physical health. For example, people with hearing difficulties may not be able to formulate accurate answers to questions if their hearing prevents them from fully understanding their wording or response options, while visually impaired people may require additional help when being presented with showcards or other visual prompts (Esposito and Jobe 1991). Drawing on the information processing framework, we argue that poor mental health and the presence of certain mental conditions may also affect the ways in which respondents provide responses to survey questions. In particular, symptoms associated with poor mental health or mental conditions may have the potential to alter the survey response process in ways that result in suboptimal survey interview outcomes. While there is variation in the nature and severity of mental health issues, 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. 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, and Mitrou 2015). In addition, there is a social stigma against people who have poor mental health (World Health Organization 2010), which may make individuals with mental health issues less open to fully engage in survey interviews due to perceived stigma and power imbalances. This suggests that, when faced with such an unfamiliar situation, individuals with these symptoms may be more likely to be apprehensive of or mistrust 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. 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 by clinical staff in establishing successful interpersonal communication and cooperation strategies with hospital patients with mental health issues (Treloar 2009, Eren and Şahin 2016).

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). Individuals who suffer mood (affective) disorders usually display symptoms characterized by depression, apathy or anhedonia, have comparatively low energy and high fatigue, reduced problem-solving capabilities, and a reduced ability to concentrate. 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 and ultimately increase response errors. Low energy and high fatigue 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 efforts in answering questions, or satisficing (Krosnick 1989). This would apply to individuals with depressive symptoms not only due to the general emotional symptoms associated with their condition, 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 face unfamiliar interviewer behaviours that make them uncomfortable (Esposito and 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 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 may be characterized by comparatively low levels of engagement and cooperation.

Third, poor mental health often displays comorbidity with *reduced faculties in cognitive capabilities* which are 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). Some mental disorders are in fact defined in terms of such cognitive difficulties, e.g. dyslexia, attention-deficit disorders or mental retardation/intellectual disability. Others, such as depression, involve temporary cognitive dysfunctions. Dementia –the most prevalent umbrella mental conditions in elderly populations in developed countries, 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 (Patten 2003, Kruijshaar et al. 2005), 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, and Hox 2000, Sigelman et al. 1980, Meisenberg and Williams 2008, Hartley and MacLean 2006).

Collectively, these general principles lead us to hypothesize that poor mental health will be associated with lower IRQSI. 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 and IRQSI using data from the HILDA Survey (Watson and 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 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 Summerfield et al. 2015 for details), and is largely representative of the Australian population in 2001. 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 household are asked to participate in the survey. 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 if 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 conditions is 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 data on IRQSI, (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 Measures of the quality of 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 IRQSI aspects, whether or not the interviewer considered that: (i) the respondent was suspicious of the study after the interview, (ii) the respondent had issues understanding the survey questions, and (iii) the respondent was not cooperative during the interview.

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

Table 1. Sample descriptive statistics

	Observations	Mean/%	SD	Minimum	Maximum
Outcome variables					
Interviewer assessment: Respondent was suspicious of the study after the interview	200,237	2%		0	1
Interviewer assessment: Respondent had issues understanding the survey questions	200,238	4%		0	1
Interviewer assessment: Respondent was not cooperative during the interview	200,239	2%		0	1
Key explanatory variables					
SF-36 Mental Health Inventory	178,252	74.20	17.14	0	100
Kessler 10 Psychological Distress Scale	53,238	15.72	6.30	10	50
Respondent has a mental illness requiring help/supervision	173,301	2%		0	1
Respondent has difficulty learning/understanding things	173,301	1%		0	1
Respondent has a nervous/emotional condition requiring treatment	173,301	4%		0	1
Control variables					
Female	200,311	53%		0	1
Age in years	200,311	44.00	18.56	14	101
Partnered	200,197	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		76%		0	1
Australian born, Indigenous		2%		0	1
Migrant from English-speaking background		10%		0	1
Migrant from non-English-speaking background		12%		0	1
Highest educational qualification	200,201				
Degree or higher degree		21%		0	1
Professional qualification		28%		0	1
School year 12		15%		0	1
Below school year 12		35%		0	1
Employment status	200,311				

Employed (including self-employment)		63%		0	1
Not in the labour force		33%		0	1
Unemployed		4%		0	1
Annual household disposable income (in $10,000s)		8.59	6.47	0	201
Area remoteness	200,311				
Major city		62%		0	1
Inner regional area		24%		0	1
Outer regional, remote or very remote area		13%		0	1
Socio-Economic Index for Areas	200,265				
1st quintile		20%		0	1
2nd quintile		20%		0	1
3rd quintile		20%		0	1
4th quintile		20%		0	1
5th quintile		20%		0	1
State of residence	200,311				
New South Wales		30%		0	1
Victoria		25%		0	1
Queensland		21%		0	1
South Australia		9%		0	1
Western Australia		9%		0	1
Tasmania		3%		0	1
Northern Territory		1%		0	1
Australian Capital Territory		2%		0	1
Number of times previously interviewed	200,311	5.81	3.91	1	14
First contact with interviewer	200,311	51%		0	1
Interviewer workload	200,311	123.78	57.61	1	389
Survey year	200,311	2008	4.13	2001	2014

Notes: HILDA Survey data, Australia, 2001-2014.

3.3 Measures of mental health and mental disorders

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 will relate to better IRQSI.

Our first mental health measure is the SF-36 Mental Health Inventory (MHI-5) (Ware and 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 outcome) to 100 (best outcome). 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 (best outcome) to 50 (worst outcome). 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.

Using responses from a HILDA Survey multi-response question available in waves 3-14 (2003-2014), we construct three additional binary variables capturing long-lasting mental-health 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 conditions that relate to mental health: (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 simple logistic regression models without control variables on each of the three outcome variables. These unadjusted models give the ‘raw’ associations between respondent mental health and IRQSI, and take the form:

$$\log\left(\frac{\Pr(Q_{ijt}=1)}{1-\Pr(Q_{ijt}=1)}\right) = \alpha + \beta_1 H_{ijt} + e_{ijt} \tag{1}$$

where the subscripts i, j and t refer to individual, interviewer, and time period, respectively; Q is a dichotomous outcome variable capturing an aspect of IRQSI, α is the model’s grand intercept; H is a given measure of respondents’ mental health and β_1 its 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 capturing IRQSI at different levels of the explanatory variables capturing mental health and disorders. These are helpful to determine the magnitude of the differences in IRQSI across individuals with different mental health levels. 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.

It is possible that the associations between the summary mental health and IRQSI are more pronounced in the low-health tail of the mental health summary measures. That is, IRQSI may only be low, or disproportionately low, amongst individuals with very low mental health scores. To assess this, we fit another set of models allowing for non-linear associations between the two continuous mental health measures (MHI-5 and K10) and the IRQSI outcome variables. This is accomplished by adding quadratic and cubic terms for the mental health variables, as follows:

$$\log\left(\frac{\Pr(Q_{ijt}=1)}{1-\Pr(Q_{ijt}=1)}\right)=\alpha+\beta_1 H_{ijt}+\beta_2 H^2_{ijt}+e_{ijt} \tag{2}$$

$$\log\left(\frac{\Pr(Q_{ijt}=1)}{1-\Pr(Q_{ijt}=1)}\right)=\alpha+\beta_1 H_{ijt}+\beta_2 H^2_{ijt}+\beta_3 H^3_{ijt}+e_{ijt} \tag{3}$$

We take statistically significant effects on the β_2 parameter in the quadratic model, and the β_3 parameter in the cubic model as evidence of non-linear relationships.

We then estimate a third set of models to test whether the associations between mental health and IRQSI 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 IRQSI are indeed due to respondents' mental health and conditions. However, we acknowledge that identifying causal relationships may not be possible with these observational data for reasons discussed below.

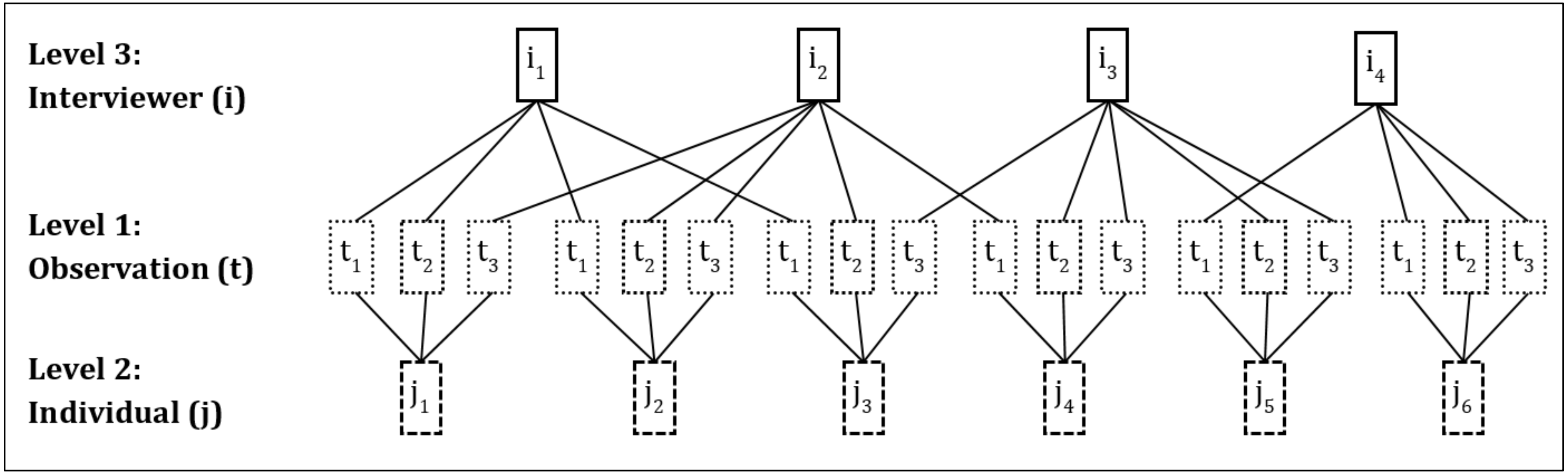
Accounting for unobserved confounders is particularly important, as a degree of subjectivity is involved in interviewers' IRQSI reports. To accomplish this, we deploy three-level (multilevel) models, as these are the optimal way to model data in which person-year observations (Level 1) are nested within survey respondents (Level 2), who are in turn nested within survey interviewers (Level 3) (Lynn, Kaminska, and Goldstein 2014, Vassallo et al. 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 (Figure 1) (Hill and Goldstein 1998, Browne, Goldstein, and Rasbash 2001). Since the outcome variables are dichotomous, we estimate logistic regression models:

$$\log\left(\frac{\Pr(Q_{ijt}=1)}{1-\Pr(Q_{ijt}=1)}\right)=\alpha+\beta_1 H_{ijt}+\gamma X_{ijt}+\sum_{t}^{T} w_{ijt}u_{jt}+v_{ij}+e_{ijt} \tag{4}$$

Here, the X_{ijt} is a vector of control variables 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, Brunton-Smith, Sturgis, and Williams 2012). The models were estimated using MLwiN 2.25 software and Markov Chain Monte Carlo (MCMC) methods (Rasbash and Browne 2008). For ease of interpretation, we report the estimates of all logistic regression models as odds ratios. Because of the complexity of these models and the data structure, the data cannot weighted for attrition or sample selection, and so the results must be interpreted with caution.

The X_{ijt} vector of control variables includes a comprehensive set of factors suspected to confound the associations between respondents' mental health and IRQSI (Table 1). These include: respondents' gender, age and its square, partnership status, number of adults in the household, number of children in the household, ethno-migrant background (Australian born; Indigenous Australian; migrant from English-speaking background; migrant from non-English-speaking background), highest educational qualification (below year 12; year 12; professional qualification; degree or higher), annual household income, area remoteness (major city; inner regional; outer regional, remote or very remote), Socio-Economic Index For Areas (quintiles), state (New South Wales; Victoria; Queensland; South Australia; Western Australia; Tasmania; Northern Territory; Australian Capital Territory), number of times interviewed, first contact with interviewer, interviewer workload, and survey year. The next section presents our empirical results.

Figure 1. Data structure, example



4 Respondent mental health and interviewer reports of survey interview quality

4.1 Unadjusted logistic regression models

Results from our unadjusted logistic regression models of IRQSI using the HILDA Survey data are summarized in Table 2, and expressed as odds ratios (OR). Results in Column 1 indicate that 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. None of the three mental condition measures is statistically significantly related to this outcome.

Results in Column 2 indicate that 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. Having a mental illness requiring help/supervision (OR=3.577, $p<0.001$), having difficulty learning/understanding things (OR=11.339, $p<0.001$), and having a nervous/emotional condition (OR=2.115, $p<0.001$) significantly increase such likelihood.

Results in Column 3 indicate that lower scores in the MHI-5 (OR=0.988, $p<0.001$) and higher K10 scores (OR=1.025, $p<0.001$) are related to interviewer reports of uncooperativeness amongst survey respondents. The same applies to having a mental illness requiring help/supervision (OR=2.264, $p<0.001$), having difficulty learning/understanding things (OR=3.319, $p<0.001$), and having a nervous/emotional condition (OR=1.320, $p<0.01$).

Table 2. Unadjusted logistic regression models of the quality of the survey interview, odds ratios

	Interviewer assessment		
	Suspicious of interview	Poor question understanding	Lack of cooperation
	(1)	(2)	(3)
Summary measures			
SF-36 Mental Health Inventory†	0.995***	0.981***	0.988***
N (observations)=178,210 / N (individuals)=27,192 / N (interviewers)=632			
Kessler 10 Psychological Distress Scale‡	1.014*	1.059***	1.025***
N (observations)=53,227 / N (individuals)=20,164 / N (interviewers)=360			
Health conditions			
Respondent has mental illness that requires help/supervision§	1.082	3.577***	2.264***
N (observations)=173,242 / N (individuals)=26,475 / N (interviewers)=556			
Respondent has difficulty learning/understanding things§	1.100	11.339***	3.139***
N (observations)=172,962 / N (individuals)=26,445 / N (interviewers)=556			
Respondent has nervous/emotional condition that requires treatment§	1.025	2.115***	1.320**
N (observations)=172,962 / N (individuals)=26,445 / N (interviewers)=556			

Notes: HILDA Survey data, Australia. † Data for years 2001-2014; ‡ Data for years 2007, 2009, 2011 & 2013; § Data for years 2003-2014. Statistical significance: * p<0.05, ** p<0.01, *** p<0.001.

4.2 Predicted probabilities

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

The magnitude of association between the mental health and disorder variables and the outcome variable capturing being suspicious of the study is very small. To illustrate this, 1.9% of individuals in the 10^{th} percentile of the MHI-5 distribution are predicted to be rated by interviewers as being suspicious of the study, compared to 1.6% of individuals in the 90^{th} percentile of the MHI-5 distribution.

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 health conditions 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%).

Effect sizes are greatest on the outcome variable capturing interviewer reports of poor question comprehension amongst respondents. For example, 4.6% of individuals in the 10^{th} 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 90^{th} percentile of the MHI-5 distribution. Amongst the health conditions, results are striking: 3.6% to 3.8% of respondents without health conditions 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.

Table 3. Predicted probabilities from unadjusted logistic regression models of the quality of the survey interview

	Percentile					Condition	
	10th	25th	50th	75th	90th	0	1
Interviewer assessment: Respondent was suspicious of interview							
SF-36 Mental Health Inventory†	1.9%	1.8%	1.7%	1.6%	1.6%		
Kessler 10 Psychological Distress Scale‡	1.1%	1.1%	1.2%	1.2%	1.4%		
Mental illness requiring help§						1.6%	1.8%
Difficulty learning/understanding§						1.6%	1.8%
Nervous/emotional condition§						1.6%	1.7%
Interviewer assessment: Respondent displayed poor question understanding							
SF-36 Mental Health Inventory†	4.6%	3.6%	2.7%	2.3%	2.1%		
Kessler 10 Psychological Distress Scale§	2.0%	2.1%	2.5%	3.1%	4.4%		
Mental illness requiring help§						3.8%	12.5%
Difficulty learning/understanding§						3.8%	7.8%
Nervous/emotional condition§						3.6%	30.0%
Interviewer assessment: Respondent was uncooperative							
SF-36 Mental Health Inventory†	1.6%	1.4%	1.2%	1.0%	1.0%		
Kessler 10 Psychological Distress Scale‡	0.9%	0.9%	1.0%	1.1%	1.2%		
Mental illness requiring help§						1.5%	3.3%
Difficulty learning/understanding§						1.5%	4.5%
Nervous/emotional condition§						1.5%	2.0%

Notes: HILDA Survey data, Australia. † Data for years 2001-2014; ‡ Data for years 2007, 2009, 2011 & 2013; § Data for years 2003-2014. Statistical significance: * p<0.05, ** p<0.01, *** p<0.001.

4.3 Non-linear associations

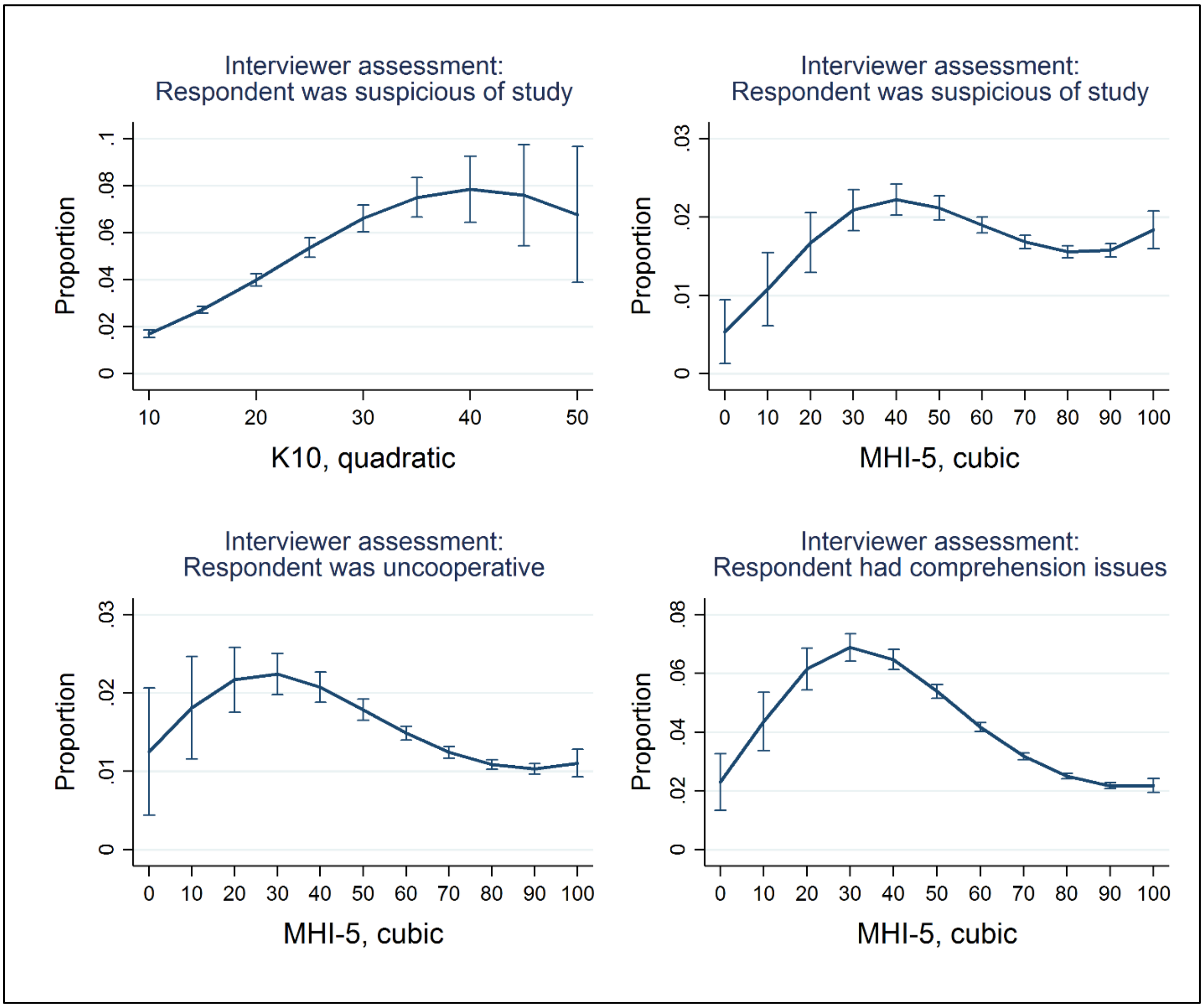
We also test for non-linear associations between the two continuous mental health measures (MHI-5 and K10) and the IRQSI outcomes variables by adding quadratic and cubic terms of the mental health measures to the unadjusted logit models discussed before. This helps determine whether or not the associations between these mental health summary variables and IRQSI concentrate on certain parts of their distribution. We find evidence of statistically significant non-linear relationships for some of the models, for which we plot the predicted probabilities across the distribution of the mental health variables in Figure 2.

The graph on the top left of Figure 2 shows predictions from a quadratic model for the K10 explanatory variable and the outcome variable capturing whether the interviewer considered that the respondent was suspicious of the study. The predicted probability of the interviewer assessing a respondent as being suspicious of the study increases with psychological distress, but at a declining rate.

The graph on the top right of Figure 2 shows results from a cubic model for the MHI-5 explanatory variable and the outcome variable capturing suspicions. Unexpectedly, very poor mental health is associated with very low levels of suspicion. However, between MHI-5 scores of 40 to 80, where most respondents fall, suspicions decrease slightly with mental health.

The two graphs at the bottom of Figure 2 show predictions from cubic models on interviewer-reported respondent uncooperativeness (left) and poor question understanding (right) using the MHI-5 explanatory variable. In both, the predicted probabilities have inverted U shapes: at low mental-health levels increasing mental health leads to worse IRQSI, while at high mental-health levels (where most respondents fall) increasing mental health leads to better IRQSI. That is, in these models the worst IRQSI is observed for individuals with ‘moderately bad’ rather than ‘extremely bad’ mental health.

Figure 2. Predicted probabilities from unadjusted logistic regression models, non-linear effects



Notes: HILDA Survey data, Australia. All of the plots are based on statistically significant associations in the models.

4.4 Three-level, cross-classified logistic regression models

Results from our three-level, cross-classified models of IRQSI using the HILDA Survey data are summarized in Table 4, and expressed as odds ratios (OR). These are revealing as to whether or not the mental health and disorder measures are associated with IRQSI when adjusting for observable and unobservable confounders. We note however that the linear relationships we estimate here will be attenuated for those models for which non-linear associations were previously reported, and that direct comparisons of odds ratios between adjusted and unadjusted logit models are inappropriate due to the scaling problem (Mood 2010).

Results in Column 1 are for models considering interviewer reports of respondents being suspicious of the study as the outcome variable. In these models, the ORs on the MHI-5 measure (OR=0.995, $p<0.001$) are statistically significant, and of a similar magnitude as those reported for the unadjusted logistic regression models. The ORs on the K10 are no longer statistically significant.

Results in Column 2 are for models in which the outcome variable captures interviewer reports of respondents experiencing issues understanding the survey questions. The ORs on the explanatory variables in these models are similar to those in the unadjusted logistic regression models. This applies to the MHI-5 (OR=0.986, $p<0.001$) and the K10 (OR=1.054, $p<0.001$) indices, as well as to the binary measures for having a mental illness requiring help/supervision (OR=2.999, $p<0.001$), having difficulty learning/understanding things (OR=7.194, $p<0.001$), and having a nervous/emotional condition (OR=1.680, $p<0.001$). The magnitude of the effects is nevertheless smaller for the mental health conditions.

Finally, results in Column 3 are for models in which the outcome is whether or not the interviewer reported that the respondent was uncooperative during the interview. These are again similar to the analogous results in Table 2 for the MHI-5 (OR=0.988, $p<0.001$) and K10 (OR=1.018, $p<0.05$) measures, and for having a mental illness requiring help/supervision (OR=2.018, $p<0.001$), having difficulty learning/understanding things (OR=2.358, $p<0.001$), and having a nervous/emotional condition (OR=1.284, $p<0.05$). Again, the ORs on the health conditions are smaller in these more complex models accounting for observable and unobservable factors.

5. Discussion and conclusion

5.1 Summary of study aims and findings

Despite increasing Government expenditure in mental health services, the prevalence of mental illness in countries such as Australia has remained stable or even increased, with one in five Australians currently suffering from a mental condition (Department of Health and Ageing 2013). In this context, gaining a robust understanding of the predictors and consequences of ill mental health 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 about the relative accuracy of the survey information gathered from individuals with poorer and better mental health.

Table 4. Cross-classified multilevel logistic regression models of the quality of the survey interview, odds ratios

	Interviewer assessment		
	Suspicious of interview	Poor question understanding	Lack of cooperation
	(1)	(2)	(3)
Summary measures			
SF-36 Mental Health Inventory†	0.995***	0.986***	0.988***
N (observations)=177,973 / N (individuals)=27,165 / N (interviewers)=632			
Kessler 10 Psychological Distress Scale‡	1.011	1.054***	1.018*
N (observations)=53,145 / N (individuals)=20,140 / N (interviewers)=360			
Health conditions			
Respondent has mental illness that requires help/supervision§	1.052	2.999***	2.018***
N (observations)=172,962 / N (individuals)=26,445 / N (interviewers)=556			
Respondent has difficulty learning/understanding things§	1.161	7.194***	2.358***
N (observations)=172,962 / N (individuals)=26,445 / N (interviewers)=556			
Respondent has nervous/emotional condition that requires treatment§	1.029	1.680***	1.284*
N (observations)=172,962 / N (individuals)=26,445 / N (interviewers)=556			

Notes: HILDA Survey data, Australia. † Data for years 2001-2014; ‡ Data for years 2007, 2009, 2011 & 2013; § Data for years 2003-2014. Controls: respondents' gender, age and its square, partnership status, number of adults in the household, number of children in the household, ethno-migrant background (Australian born; Indigenous Australian; migrant from English-speaking background; migrant from non-English-speaking background), highest educational qualification (below year 12; year 12; professional qualification; degree or higher), annual household income, area remoteness (major city; inner regional; outer regional, remote or very remote), Socio-Economic Index For Areas (quintiles), state (New South Wales; Victoria; Queensland; South Australia; Western Australia; Tasmania; Northern Territory; Australian Capital Territory), number of times interviewed, first contact with interviewer, interviewer workload, and survey year. Full tables of coefficients are available from the authors upon request. Statistical significance: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

In this paper, we contributed to filling this gap in knowledge form the prism of interviewer observations. Drawing on information processing theory, we hypothesized that individuals with low levels of mental health and with mental conditions would display lower IRQSI due to comparatively low cognitive and motivational processing in answering survey questions, emerging from 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. In our empirical analyses, we tested how interviewer reports of the quality of the survey interview were related to the mental health of survey respondents, using a unique panel dataset that is largely representative of the Australian population and state-of-the-art multilevel regression models.

Our findings are consistent with the expectations outlined before: the mental health of survey participants is related to IRQSI and individuals with poorer mental health are more likely to display low IRQSI. These associations were visible across a range of IRQSI outcomes (interviewers reporting that respondents were suspicious of the study, had issues understanding survey questions, and were uncooperative) and measures of mental health and disorders (the MHI-5, the K10, and three binary indicators of long-lasting mental health conditions).

However, the magnitude of the associations varied across models. Differences in IRQSI by mental health were more pronounced and more often statistically significant for the outcome variables measuring interviewer ratings of respondent cooperation and question comprehension than for the outcome variable measuring interviewer ratings of respondent suspicions. They were also visibly larger for the measures capturing mental health conditions than the summary mental health measures. Some non-linear associations were also reported for the summary mental health conditions, but they did not show a consistent pattern.

Statistically significant associations between the measures of mental health and conditions and the IRQSI outcome variables are also apparent in multivariate logistic regression models accounting for observable and unobservable 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 IRQSI amongst respondents with poor mental health constitute new and important 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 and Wooden 2009). The lower IRQSI observed amongst individuals with poor mental health has important implications for how researchers undertake survey research on mental health and how they interpret the results. To the extent that professionally-trained interviewers are accurate in their assessments, this finding is suggestive that the accuracy of the resulting survey data is comparatively lower amongst respondents with poor mental health. Hence, it is possible that survey analyses of individuals with poor mental health produce unreliable results – both when comparisons are made between these individuals and individuals with better mental health, and when the (sub)populations of interest comprise a large fraction of respondents with poor mental health. This would pose a significant challenge to the usefulness of findings generated using survey data to inform the design of evidence-based mental health policy.

In principle, there are two ways in which these issues could be addressed or ameliorated. A first way is for researchers to devise and implement statistical solutions that minimize any errors or biases in the survey information collected from individuals with poor mental health. At a basic level, one can explicitly control for the IRQSI variables in regression models (see e.g. Peytchev and Peytcheva 2007) and examine whether doing so changes the estimated relationships of interest. More powerful approaches might involve techniques that more directly incorporate the associated measurement error in the statistical models (Buonaccorsi, 2010). These have already been successfully implemented in cognate fields of inquiry, e.g. in cross-cultural survey research (King et al. 2004).

A second and more costly way to account for differential survey quality by mental health is to reconsider how individuals with poor mental health engage with the survey process. 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, then the survey instruments, study protocols and interview setting could be adapted to optimize IRQSI outcomes. Survey practitioners could also provide some basic training to survey interviewers on how to maximize data quality from respondents with low mental health (Becker et al. 2004). This is similar to the cultural competence training that is sometimes provided to survey interviewers, as well as public sector employees such as health professionals, who frequently work with individuals from vulnerable populations such as ethnic minorities and LGBT people (Mays 2001, Betancourt et al. 2003, Westerman 2004); or the training provided to lecturers and other staff at Higher Education institutions on dealing with people with mental health issues.

Addressing these data shortcomings is particularly relevant for surveys aimed explicitly at gathering information on individuals with mental health issues (e.g. medical expenditure surveys), or surveys focused on population subgroups in which such issues are relatively prevalent (e.g. elderly people, crime victims, war veterans, or sexual minorities). Studies involving cognitive interviewing techniques or detailed examinations of interviewer-interviewee interactions could be designed to shed light over how survey processes can be tailored to better address the needs of these individuals (Hartley and MacLean 2006).

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 methodological refinement. First, individuals with poor mental health and disorders are less likely to participate in surveys, remain within the sample of panel surveys, and complete and return self-completion questionnaires (such as the one containing the summary mental health measures within the HILDA Survey). In addition, the HILDA Survey sample does not include 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 issues. 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 disorders on IRQSI 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 conditions 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) (Kessler and Ustun 2004). They are also very coarse, failing to reflect the complexity of mental disorders reflected in the International Classification of Diseases (ICD-10) or the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). As a result, the broad results that we present here may mask substantial heterogeneity and may differ when other measurement tools for mental conditions are employed. Further research using alternative measures of mental conditions is warranted.

Third, we do not claim that the associations we find are causal. In fact, some of the estimated effect of respondent mental health on IRQSI 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 interview survey quality when they engage with 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 low survey quality assessments due their own prejudice. In fact, interviewers may be aware of the respondents' mental health and conditions through their knowledge of respondents' survey answers. In this respect, while the summary mental health measures in the HILDA Survey are completed privately, the information on health conditions is gathered in the face-to-face survey interview. It is difficult to imagine ways to accurately correct for this source of reverse causation using observational data with no information on interviewers' attitudes to mental health issues. 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 of survey respondents.

Finally, there is a surprising paucity of evidence on the degree to which interviewer reports of survey quality, such as those employed in this study, actually correlate with objective measures of data quality (beyond some evidence linking them to attrition in panel studies). Hence, future studies may complement our findings by additionally considering how respondents' mental health and conditions are associated with other indicators of survey data quality which are not reported by interviewers. These may include the proportion of survey items to which the respondent refused to provide an answer or to which the respondent provided an implausible or 'don't know' answer, and the prevalence of unusually short, long and interrupted survey interviews.

5.4 Concluding remarks

While surveys are powerful means by which to gather evidence to inform the development of health policies, it is not clear that researchers and policymakers should take the accuracy of survey data generated from respondents with ill mental health for granted. More research aimed at comparing how individuals with poorer and better mental health engage in the survey process, and whether and how their poor mental health is related to the quality of the information retrieved from these individuals is sorely needed.

References

Aneshensel, Carol S., Jo C. Phelan, and Alex Bierman. 2013. *Handbook of the Sociology of Mental Health*. 2nd ed. New York: Springer.

Australian Bureau of Statistics. 2009. *4327.0 - National Survey of Mental Health and Wellbeing: Users' guide, 2007*. Canberra: Australian Bureau of Statistics.

Becker, Heather, Greg Roberts, Janet Morrison, and Julie Silver. 2004. "Recruiting people with disabilities as research participants: Challenges and strategies to address them." *Mental Retardation* 42 (6):471-475.

Betancourt, Joseph R., Alexander R. Green, J. Emilio Carrillo, and Owusu Ananeh-Firempong. 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, Natacha, Edith de Leeuw, and Joop Hox. 2000. "Children as Respondents in Survey Research: Cognitive Development and Response Quality 1." *Bulletin de Méthodologie Sociologique* 66 (1):60-75.

Browne, William J, Harvey Goldstein, and Jon Rasbash. 2001. "Multiple membership multiple classification (MMMC) models." *Statistical Modelling* 1 (2):103-124.

Brunton-Smith, Ian, Patrick Sturgis, and Joel Williams. 2012. "Is Success in Obtaining Contact and Cooperation Correlated With the Magnitude of Interviewer Variance?" *Public Opinion Quarterly* 76 (2):265-286.

Buonaccorsi, John P. 2010. *Measurement error: Models, methods, and applications*. New York: Chapman & Hall.

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.

Dijkstra-Kersten, Sandra M A, Karolien E M Biesheuvel-Leliefeld, Johannes C van der Wouden, Brenda W J H Penninx, and Harm W J van Marwijk. 2015. "Associations of financial strain and income with depressive and anxiety disorders." *Journal of Epidemiology and Community Health* 69 (7):660-665.

Durrant, Gabriele B., Robert M. Groves, Laura Staetsky, and Fiona Steele. 2010. "Effects of interviewer attitudes and behaviours on refusal in household surveys." *Public Opinion Quarterly* 74 (1):1-36.

Eren, N., and S. Şahin. 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.

Esposito, James L., & Jobe, Jared B. 1991. "A general model of the survey interaction process". In *Bureau of the Census Seventh Annual Research Conference Proceedings. Washington DC: US Bureau of the Census* (pp. 537-560).

Groves, Robert M. 2004. *Survey Errors and Survey Costs*. Hoboken: John Wiley & Sons, Inc.

Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. 2nd ed.. ed. Hoboken, N.J.: Wiley.

Hartley, S. L., and W. E. MacLean. 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.

Hewitt, Belinda, Gavin Turrell, and Katrina Giskes. 2012. "Marital loss, mental health and the role of perceived social support: findings from six waves of an Australian population based panel study." *Journal of Epidemiology and Community Health* 66 (4):308-314.

Hill, Peter W., and Harvey Goldstein. 1998. "Multilevel modeling of educational data with cross-classification and missing identification for units." *Journal of Educational and Behavioral Statistics* 23 (2):117-128.

Johnston, David, Carol Propper, Stephen Pudney, and Michael Shields. 2014. "Child mental health and educational attainment: multiple observers and the measurement error problem." *Journal of Applied Econometrics* 29 (6):880-900.

Jones-Rounds, McKenzie L, Gary W Evans, and Matthias Braubach. 2014. "The interactive effects of housing and neighbourhood quality on psychological well-being." *Journal of Epidemiology and Community Health* 68 (2):171-175.

Kennickell, Arthur B., Martha Starr-McCluer, and Brian J. Surette. 2000. "Recent changes in U.S. family finances: Results from the 1998 Survey of Consumer Finances." *Federal Reserve Bulletin* 86 (1):1-29.

Kessler R. C. G., and Ustun T.B. 2004. The World Mental Health (WHM) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research* 13 (2):93-121.

Kessler, R. C., G. Andrews, L. J. Colpe, E. Hiripi, D. K. Mroczek, S. L. Normand, E. E. Walters, and A. M. Zaslavsky. 2002. "Short screening scales to monitor population prevalences and trends in non-specific psychological distress." *Psychological Medicine* 32 (06):959-976.

King, G., C. J. Murray, J. A. Salomon, and A. Tandon. 2004. "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review* 98 (01):191-207.

Koenen, K. C., T. E. Moffitt, A. L. Roberts, L. T. Martin, L. Kubzansky, H. Harrington, R. Poulton, and A. Caspi. 2009. "Childhood IQ and adult mental disorders: a test of the cognitive reserve hypothesis." *American Journal of Psychiatry* 166 (1):50-7.

Krosnick, Jon A. 1991. "Response strategies for coping with the cognitive demands of attitude measures in surveys." *Applied Cognitive Psychology* 5 (3):213-236.

Kruijshaar, Elisabeth Michelle, Jan Barendregt, Theo Vos, Ron de Graaf, Jan Spijker, and Gavin Andrews. 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.

Lahelma, Eero, Tea Lallukka, Mikko Laaksonen, Peppiina Saastamoinen, and Ossi Rahkonen. 2012. "Workplace bullying and common mental disorders: a follow-up study." *Journal of Epidemiology and Community Health* 66 (6):e3.

Lynn, P., O. Kaminska, and H. Goldstein. 2014. "Panel attrition: How important is interviewer continuity?" *Journal of Official Statistics* 30 (3):443-457.

Mays, Vicky M. 2001. "Methods for increasing recruitment and retention of ethnic minorities in health research through addressing ethical concerns." *Proceedings of the Seventh Conference on Health Survey Research Methodology*. Washington, D.C.: National Center for Health Statistics, USDHHS.

Meisenberg, Gerhard, and Amandy Williams. 2008. "Are acquiescent and extreme response styles related to low intelligence and education?" *Personality and Individual Differences* 44 (7):1539-1550.

Mood, Carina. 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.

Patten, S. B. 2003. "Recall bias and major depression lifetime prevalence." *Social Psychiatry and Psychiatric Epidemiology* 38 (6):290-6.

Perales, F., B. Baffour, and F. Mitrou. 2015. "Ethnic differences in the quality of the interview process and implications for survey analysis: The case of Indigenous Australians." *PLoS ONE* 10 (6):e0130994.

Perales, F., J. del Pozo-Cruz, and B. del Pozo-Cruz. 2014. "Impact of physical activity on psychological distress: A prospective analysis of an australian national sample." *American Journal of Public Health* 104 (12):e91-97.

Peytchev, Andy, and Emilia Peytcheva. 2007. "Relationship between measurement error and unit nonresponse in household surveys: An approach in the absence of validation data." Joint Statistical Meeting, Salt Lake City, UT.

Power, A. Kathryn. 2010. "Transforming the nation's health: Next steps in mental health promotion." *American Journal of Public Health* 100 (12):2343-2346.

Rasbash, Jon, and William J. Browne. 2008. "Non-hierarchical multilevel models." In *Handbook of Multilevel Analysis*, edited by Jan de Leeuw and Erik Meijer, 301-334. New York: Springer.

Rudolph, Kara E, and William W Eaton. 2016. "Previous anxiety and depression as risk factors for early labour force exit." *Journal of Epidemiology and Community Health* 70 (4):390-395.

Schwarz, Norbert. 2007. Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, 21 (2):277-287.

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

Summerfield, Michelle, Simon Freidin, Markus Hahn, Ning Li, Ninette Macalalad, Laura Mundy, Nicole Watson, Roger Wilkins, and Mark Wooden. 2015. *HILDA User Manual – Release 14.*

Melbourne: Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Taylor, Marcia Freed, John Brice, Nick Buck, and Elaine Prentice-Lane. 2010. *British Household Panel Survey User Manual Volume A: Introduction.* Colchester: University of Essex.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

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

Tzoumakis, Stacy, Patrick Lussier, and Raymond R. Corrado. 2015. "Profiles of maternal parenting practices: Exploring the link with maternal delinquency, offending, mental health, and children's physical aggression." *International Journal of Offender Therapy and Comparative Criminology* 59 (12):1267-1296.

Vassallo, Rebecca, Gabriele B. Durrant, Peter W. F. Smith, and Harvey Goldstein. 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.

Ware, John E., and Cathy Donald Sherbourne. 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., and M. Wooden. 2009. "Identifying factors affecting longitudinal survey response." In *Methodology of Longitudinal Surveys*, edited by P. Lynn, 157-181. Chichester: John Wiley and Sons.

Watson, Nicole, and Mark Wooden. 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.

Westerman, Tracy 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.