Contents lists available at ScienceDirect

Social Science & Medicine

Who provides inconsistent reports of their health status? The importance of age, cognitive ability and socioeconomic status☆

Nicole Black [a], David W. Johnston [a], Michael A. Shields [a], Agne Suziedelyte [b, *]

[a] Centre for Health Economics, Monash University, Australia
[b] Department of Economics, City, University of London, UK

A R T I C L E   I N F O

A B S T R A C T

Self-assessed health (SAH) measures are widely used in models of health and health inequalities. Such models assume that SAH is a reliable measure of health status. We utilise a unique feature of a national longitudinal survey to examine the consistency of responses to a standard SAH question that is asked twice to the same individual in close temporal proximity in up to three waves (2001, 2009, and 2013). In particular, we analyse whether the consistency of responses varies with personal characteristics. The main analysis sample includes 18,834 individual-year observations. We find that 57% of respondents provide inconsistent reports at least once. Characteristics that are associated with significantly higher inconsistencies are age, education, cognitive ability, and time between responses. The results suggest that there are systematic differences in the ability of individuals' to self-evaluate and summarise their own health. Consequently, failure to account for such error may lead to large estimation biases in models of health outcomes, particularly with respect to the relationship between education, cognitive ability, and health.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Every year numerous published articles in epidemiology, public health and economics use self-assessed health (SAH) to provide new knowledge about the demographic and socioeconomic characteristics associated with good health and the extent of health inequalities. Such evidence is then used to motivate health and health care policy. In particular, a large literature has examined the relationship between self-assessed health and education, unemployment, household income, occupation, wealth, neighbourhood deprivation, early life circumstances and retirement (see, for example, Meer et al., 2003; Contoyannis et al., 2004; Kunst et al., 2005; Cutler and Lleras-Muney, 2010; Johnson, 2010; and Hu et al., 2016). In this study we examine the reliability of general

SAH, arguably the most commonly analysed measure of health, where respondents are asked to rate their current general health on an ordinal scale (for example, excellent, very good, good, fair, poor). If it is the case that the willingness or capability of individuals to consistently answer this type of question is associated with certain individual-level characteristics and traits (such as education and cognitive ability), then inferences about inequalities in health may be misleading.

To address this issue we take advantage of a peculiarity in a large longitudinal survey that provides a 'quasi-experiment' for analysing reporting consistency in SAH. In three waves of the Household, Income and Labour Dynamics in Australia (HILDA) survey, respondents are asked, within a short period of time, to report their general health using two different survey modes: face-to-face interview (FFI) and self-completion questionnaire (SCQ). By analysing the variation in responses, we are able to document the extent of inconsistency in SAH and determine whether the inconsistency is influenced by individuals' characteristics and traits including their education, employment status, cognitive ability, and personality.

There are several reasons why respondents may report their health differently, even if asked on the same day when their underlying level of health arguably remains unchanged. First, it is

---

* Corresponding author. Department of Economics, City, University of London, Northampton Square, London EC1V 0HB, UK.
E-mail address: agne.suziedelyte@city.ac.uk (A. Suziedelyte).

likely that a degree of uncertainty exists in identifying underlying health levels, and therefore individuals assess their health with some "error" each time (Crossley and Kennedy, 2002). Given the cognitive demands placed on respondents, such as comprehending the question, recalling information from memory, and communicating the response (Bowling, 2005), it is likely that cognitive ability will affect the consistency of responses. Similarly, certain personality traits, such as conscientiousness, may affect the effort and consideration that is taken when answering survey questions and therefore on the reliability of responses (Bertrand and Mullainathan, 2001). Second, it is conceivable that the mode of data collection influences responses (Bowling, 2005). Different modes require different skills (for example, verbal and listening in interviews, and reading and writing in paper questionnaires). Moreover, the social nature of interviews may induce individuals to give more positive and socially desirable responses, which is known as social desirability bias (Bowling, 2005). This bias can also be associated with certain individual characteristics and traits. For example, older individuals may understate any health problems in an interview so as to appear healthy and robust in front of the interviewer. In contrast, unemployed individuals may feel socially conditioned to inflate their health problems in order to help justify their unemployment status (Anderson and Burkhauser, 1985). Third, there may be "learning effects", whereby preceding questions influence or frame the respondent's perception about their health. In particular, specific health and disease questions can influence subsequent responses about general health (Bowling and Windsor, 2008).

Our study complements a growing literature that examines reporting heterogeneity in SAH across different groups of individuals. These studies, which typically condition on a measure of underlying latent health, have identified a number of reporting phenomena, including justification bias (Bound, 1991; Kerkhofs and Lindeboom, 1995; Lindeboom and Kerkhofs, 2009), reference group effects (Wiseman, 1999; Groot, 2000) and heterogeneity in the interpretation and use of response scales (Lindeboom and Van Doorslaer, 2004; Etilé and Milcent, 2006; Bago d'Uva et al., 2011). Other related studies have confirmed that survey design affects responses, and that these effects differ by certain individual characteristics (Lumsdaine and Exterkate, 2013; Holford and Pudney, 2015). However, to the best of our knowledge, few studies have had the ability to examine the extent, characteristics, and consequences of inconsistency in responses to two near identical questions asked in close temporal proximity to the same individual. The closest examples are Crossley and Kennedy (2002) and Clarke and Ryan (2006), which focus on survey mode and question ordering effects. Both studies find that around 30% of respondents changed their responses to the general health question asked in different survey modes. While Crossley and Kennedy (2002) conclude that both survey mode and question ordering are likely to play a role, Clarke and Ryan (2006) find that survey mode has the dominant role in response changes.

We build on both of these studies in a number of important ways. First, we examine a much wider range of individual characteristics to provide a more complete understanding of who reports their health inconsistently. Many of the characteristics we examine, such as cognitive ability and personality traits are key drivers of reporting behaviour, yet are rarely measured in household surveys. Second, our study has the advantage of having three waves of data, and we use statistical models that allow us to take into account repeated observations by individuals and households. This is important because inconsistencies in reporting are likely to be clustered at both the individual and household level; our approach enables this assumption to be tested for the first time. A further contribution of our study is the ability to separate out the influences of survey mode from those of question order, which has not been feasible in past studies.

2. HILDA and the quasi-experiment

Our data is drawn from the Household, Income and Labour Dynamics in Australia survey (HILDA), which is a nationally representative longitudinal study of Australian households that began in 2001. Wave 1 contained a sample of 19,914 panel members from 7682 households, and in each subsequent year household members have been followed-up, along with any new household members resulting from changes in the composition of the original household. New households were included in the wave 11 top-up sample. The household response rates range from 87.0 per cent in wave 2 to 70.8 per cent in wave 11, while the household response rates for those households responding in the previous wave ranges from 87.0 per cent in wave 2 to 96.4 per cent in wave 11 (Summerfield et al., 2012). Annual data is currently available from 2001 to 2014, and each year includes detailed information on income, employment, health and other demographic and socio-economic information.

In every wave, HILDA includes a confidential paper self-completion questionnaire (SCQ). The first question of the SCQ reads, "In general, would you say your health is", with respondents instructed to cross one box on a 5-point ordinal scale with the labels: Excellent, Very good, Good, Fair, and Poor. In waves 1, 9 and 13 (2001, 2009 and 2013) the face-to-face interview (FFI) includes a very similar SAH question: "In general, how would you rate your health? Is it excellent, very good, good, fair or poor?" Therefore, in three waves of HILDA, individuals are asked two near-identical SAH questions, with identical ordinal scales.

Aside from mode of administration, the FFI and SCQ health questions differ with respect to the preceding questions, which can have framing effects (Bertrand and Mullainathan, 2001). The FFI question is located near the end of the interview at the beginning of a health module, and after modules on education, employment, and income, while the SCQ question is the first question on the SCQ and is also followed by a series of health questions. Therefore, the type and magnitude of priming will depend upon the ordering of the FFI and SCQ, which implies that the likelihood of inconsistency may be a function of ordering. Importantly, we are able to test for this because the ordering of FFI and SCQ was not fixed; some HILDA respondents completed the FFI before the SCQ and some after. In waves 9 and 13, the respondents record the date they complete the SCQ, and so we know that 43% of the SCQs and FFIs are completed on the same day (although we do not know the ordering). Of the remaining 57%, 80% completed the FFI before the SCQ.

In all analyses, we restrict the sample to individuals who are at least 25 years old and are not studying at the time of the interview, because completed education is one of our variables of interest. Our main analysis sample includes individuals for whom we know the date of the SCQ (i.e. who are present in waves 9 and 13) and who complete the FFI and SCQ not more than 30 days apart (97.4%). This allows us to examine the influence of days between SAH questions on response consistency and ensure that the underlying health of respondents remains unchanged between questions. The median and mean days between questions are 1 and 2.5 days, respectively. We also exclude phone interviews. The main analysis sample includes 18,834 individual-year observations.

Table 1 provides a brief description and mean values of the individual characteristics and traits that we use in our statistical modelling of response inconsistency. Largely following the literature (Crossley and Kennedy, 2002; Clarke and Ryan, 2006; Lumsdaine and Exterkate, 2013), our base set of control variables are gender (47% are male), age (average age is 50.5 years), marital