



# “Your Health at Present”: Are Patterns of Reporting Heterogeneity in Self-rated Health Gendered?

Ralitza Dobreva<sup>1</sup> · Dorrit Posel<sup>2</sup>

Received: 12 September 2022 / Accepted: 15 May 2023 / Published online: 13 June 2023  
© The Author(s) 2023

## Abstract

This study explores reporting heterogeneity in self-rated health in a developing country context, and it thereby contributes to a growing literature which evaluates whether purely subjective indicators of health provide reliable measures of health status that permit inter-personal comparisons. We focus on gender and investigate if reporting patterns on self-rated health vary systematically between women and men. Most studies assess the reliability of self-rated health indicators by incorporating gender as a control variable in the analysis. We extend this research by examining whether gender intersects with a range of social factors to affect health perceptions. We analyse national longitudinal data, using the random effects generalised ordered probit iterative estimation procedure, thereby controlling for unobserved characteristics that may compromise cross-sectional analysis. Overall, our study corroborates other findings on the reliability of subjective health indicators, even in the culturally and economically diverse context of South Africa. However, although reporting styles are largely consistent between women and men, we also identify several differences, which warrant further investigation in other contexts.

**Keywords** Self-rated health · Reporting heterogeneity · Gender differences · Longitudinal data · Random effects generalised ordered probit · South Africa

JEL I10 · J16

---

✉ Ralitza Dobreva  
dobrevr@ukzn.ac.za

Dorrit Posel  
dorrit.posel@wits.ac.za

<sup>1</sup> School of Accounting, Economics and Finance, University of KwaZulu-Natal, Durban, South Africa

<sup>2</sup> School of Economics and Finance, University of the Witwatersrand, Johannesburg, South Africa

## Introduction

Health is an integral component of an individual's quality of life (Fallowfield, 2009; Khan & Raeside, 2014). Individuals who are in poor health often find it harder to participate in society, and the absence of good health makes it more difficult for people to lead fulfilling lives (Ge et al., 2019; Kaplan, 2003; Kaplan & Baron-Epel, 2003; Smith et al., 1999). However, because health is a multidimensional concept, including not only physical but also mental health and social well-being (Cutler et al., 1997), its measurement at a national scale is challenging (Ziebarth, 2010).

The simplest indicator of an individual's overall health, which can be derived from survey instruments, is their own (typically one-word) assessment in response to the question: "How would you describe your health at present?". Self-rated health (SRH) is widely used in the social and medical sciences, partly because of the ease with which it can be collected, and because of the broad scope of health that it likely reflects (Bailis et al., 2003; Idler & Benyamini, 1997). However, given its subjective nature, there is concern that the reliability and comparability of SRH across individuals or groups can be distorted by individual differences in reporting behaviour, referred to as reporting heterogeneity (Lindeboom & van Doorslaer, 2004). If these differences are systematic, then comparisons in health status across groups of individuals will be biased (Baidin et al., 2021).

This study investigates whether there is evidence of reporting heterogeneity in self-rated health in South Africa, and it thereby contributes to a growing literature which evaluates whether purely subjective indicators of health provide reliable measures of health status. Despite its extensive use internationally, there are few rigorous studies that interrogate reporting heterogeneity in SRH for developing countries (some examples include Wu et al., (2013) for China; Paul & Valtonen, (2016) for Russia; and Rossouw et al., (2018) for South Africa). Heterogeneity in reporting could be expected to be a particularly important problem for South Africa, given the country's large socio-economic disparities, the diversity of cultural and language groups, its economic geography, including the relative isolation of rural-dwellers (the majority of whom are women and children), a long history of segregation policies, and its dual healthcare system (for the poorer and uninsured, and for the richer with private health insurance).

The study focuses on gender differences in reporting patterns on self-rated health. In South Africa, as in many other countries, women report higher rates of morbidity than men, although life expectancy is lower among men (Denton & Walters, 1999; Lehohla, 2013). We probe whether there are systematic differences in reporting styles on self-rated health among women and men, and therefore whether gender comparisons of SRH in South Africa are robust.

Earlier research on South Africa, and elsewhere, has typically incorporated gender as a control variable in the analysis, and therefore has not explored whether gender intersects with other characteristics that may influence reporting behaviour (Denton & Walters, 1999). However, a range of biological, psychological, historical, and social influences may produce gender differences

in how health and its related behaviours are transformed into health perceptions (Benyamini et al., 2003; Denton & Walters, 1999; Denton et al., 2004).

To investigate reporting heterogeneity in South Africa, we analyse longitudinal data collected in the National Income Dynamics Study (NIDS). As the data are nationally representative, we can draw inferences about population differences in reporting styles. The use of a longitudinal data estimation technique – specifically, random effects generalised ordered probit iterative estimations – makes it possible to control for two kinds of heterogeneity. First, the random effects specification captures unobserved individual characteristics that may influence reporting behaviour on health, and that are not accounted for in the analysis of cross-sectional data. Second, the results reflect the observed heterogeneity in reporting styles.

In the next section, we review the literature on the reliability of SRH and problems of reporting bias. We discuss the national panel data which we analyse in Sect. 3, and outline the methodology used to interrogate the reliability of SRH. In Sect. 4, we describe the data; in Sect. 5, we present the results of the econometric analysis and in Sect. 6, we investigate whether the results may be compromised by attrition between the data waves. In the final section, we summarise and discuss the main findings of the study.

## Review: Self-rated Health (SRH) and Reporting Bias

The measurement of health presents a number of practical and conceptual difficulties because “health” is a “multiattribute concept, and not all of its attributes can be easily determined” (Cutler et al., 1997: 218). SRH offers clear practical advantages over other more “objective” measures (such as external assessments by medical professionals, or a range of questions about symptoms and diagnoses), and particularly in developing country contexts: it is easy, cheap, and fast to collect in surveys by using a single question. It also has several conceptual advantages. First, it is an all-encompassing measure: SRH is seen as “a global self-evaluation of health” (Bailis et al., 2003:203) and “while researchers are measuring the parts, respondents have access to the whole” (Idler & Benyamini, 1997:28). Second, SRH is likely to be a dynamic appraisal, including improvements, deteriorations, and elements of the respondent’s enduring self-concept (Bailis et al., 2003; Huisman & Deeg, 2010; Idler & Benyamini, 1997). Third, SRH may encapsulate the availability of physical, emotional or social resources that can either moderate a health decline, or aid a recovery after illness (Idler & Benyamini, 1997).

Evidence in favour of SRH as an effective measure of overall health is its independent ability to predict mortality, even when used *in addition to* a range of more objective health indicators and health-related behaviours – e.g. measures of chronic conditions, physical functioning, medication use, healthcare utilisation, smoking, alcohol consumption, and physical exercise, among others (Ardington & Gasealahwe, 2014; DeSalvo et al., 2006; Idler & Benyamini, 1997).

However, the subjective nature of SRH, which underpins its advantages, is also the source of its main weakness: it is affected by the mechanisms behind perception-formation and by individual measurement error (Crossley & Kennedy, 2002; Sen,

2002). Respondents' opinions are shaped by their experiences and context – for example, people who are educated and those who have access to better healthcare are likely to be better informed about their illnesses, while people in poorer settings may not perceive a burden of symptoms as out of the ordinary (Sen, 2002). Systematic variations in individuals' perceptions, reflected in different reporting styles, lead to reporting heterogeneity. In this case, SRH will not be comparable between countries, or within countries, between different groups of people (Jürges, 2007; Lindeboom & van Doorslaer, 2004; Sen, 2002).

To interrogate the reliability of SRH, an important strand of the literature investigates differences in SRH due to “true health” and those due to reporting styles or systematic patterns of measurement error (Crossley & Kennedy, 2002; Jürges, 2007; Kerkhofs & Lindeboom, 1995; Lindeboom & van Doorslaer, 2004; Pfarr et al., 2012). Assuming that the underlying “true health” variable is unobservable (latent) and continuous, then respondents answering the question about their general health project their perceptions into a reporting category from “excellent” to “poor” (Jürges, 2007; Lindeboom & van Doorslaer, 2004). When the interpretations of the category responses differ between groups – for example, if older people view health deficiencies less harshly than younger respondents with the same conditions (Groot, 2000) – then this kind of discrepancy causes a parallel shift in the entire distribution of SRH, which is referred to as an “index shift” (Lindeboom & van Doorslaer, 2004).

The threshold points between categories may also differ between groups, making one or more of the categories wider or narrower. This is often labelled a “cut-point shift” (Lindeboom & van Doorslaer, 2004). For example, cultural norms that favour an upbeat attitude may induce respondents to avoid the worst response option (Crossley & Kennedy, 2002).

Systematic reporting heterogeneity compromises inter-group comparisons of SRH. In this study, we interrogate gender differences in reporting heterogeneity in SRH, where our starting point is the paradox that men give more positive ratings to their health, yet women have lower mortality rates at every age (Case & Paxson, 2005; Denton & Walters, 1999; Idler, 2003; Kiecolt-Glaser & Newton, 2001).

Denton et al. (2004) propose two hypotheses for why women would experience higher rates of morbidity than men. The first is the differential exposure hypothesis, which argues that higher self-reports derive from women facing more limited access to resources and experiencing more stressful living circumstances, often related to gender roles. The second is the differential vulnerability hypothesis, which predicts that women have dissimilar responses to men, to the structural, behavioural and psychosocial determinants of health, and these are reflected in their higher rates of morbidity (Case & Paxson, 2005; Denton et al., 2004; Macintyre et al., 1996). If women systematically provide lower reports of self-rated health than comparable men, then this could be because women respond more negatively to the social influences on health, as predicted by the differential vulnerability hypothesis, and we would expect to see more evidence of reporting heterogeneity among women than among men.

A range of other characteristics have also been found to correlate with reporting behaviour. For example, age appears to drive reporting heterogeneity because age

influences the choice of reference group and individuals tend to become more tolerant of health conditions, or to manage them better as they age (Jylhä, 2009). Education, which can influence health via income or occupation, also modifies information processing and decision-making, making it a possible correlate of reporting heterogeneity (Cutler & Lleras-Muney, 2006). Marital status is another correlate of health perceptions. A well-functioning marriage, for example, has been found to facilitate a positive outlook, which would affect how health is assessed; while marital distress, which is strongly linked to depression (Kiecolt-Glaser & Newton, 2001; Guner et al., 2018), would negatively affect health reports. Locational factors (including country of residence, geographical area type, residence type or neighbourhood) proxy for living and working conditions, as well as access to services, and consequently, they may influence perceptions of what counts as poor, fair or excellent health (Brownson et al., 2009; Kawachi & Berkman, 2003).

To date, four studies have investigated aspects of reporting bias in self-rated health measures in South Africa (Boyce & Harris, 2011; Charasse-Pouélé & Fournier, 2006; Jelsma & Ferguson, 2004; Rossouw et al., 2018). Boyce and Harris (2011) and Charasse-Pouélé and Fournier (2006) focus on race as a proxy for socioeconomic status in South Africa, but also as a driver of differences in reporting behaviour. Jelsma and Ferguson (2004) investigate heterogeneity in self-ratings of health between respondents of different material means, proxied by household income or wealth, while Rossouw et al. (2018) consider both race and wealth, but only for elderly adults (aged 50 and above).

Whilst two of these studies (Charasse-Pouélé & Fournier, 2006; Rossouw et al., 2018) include gender when estimating health status, neither explores whether gender intersects with other characteristics. However, existing research identifies that, as in most countries, women in South Africa are more likely to suffer from chronic diseases or disabilities, which are not life-threatening, while men are more likely to suffer from conditions with a higher risk of death (Lehohla, 2013). Also in line with international patterns, a greater proportion of South African women than men experiences low-level mental health problems (Herman et al., 2009). Our study seeks to fill the gap in possible reporting bias in SRH, by interrogating gender differences in self-reports, where we consider a wide array of factors, in addition to race and socioeconomic status, that likely shape heterogeneous reporting styles, and use an estimation technique that takes advantage of the national panel data which we analyse.

## Methods and Data

### Estimating Reporting Heterogeneity: Cut-point Shifts vs Index Shifts

We investigate reporting heterogeneity using the model developed by Kerkhofs and Lindeboom (1995). The model assumes that "true health" ( $H^T$ ) is a continuous latent variable. The individual's assessment of "true health" is reflected in their response to the survey question (on SRH), represented by  $H^S$ . Since, by definition, "true health" is unobservable, more objective health measures are used as benchmarks against which to identify differences in reporting styles (Kerkhofs

& Lindeboom, 1995; Lindeboom & van Doorslaer, 2004). (These measures are discussed below.)

SRH ( $H^S$ ) can be viewed as determined by “true health” ( $H^*$ ), reporting behaviour ( $X_1$ ), and a random error component ( $\varepsilon_1$ ). The effects of reporting behaviour on SRH are captured by the  $\beta_1$ ’s.

$$H^S = f_1(H^*, X_1, \varepsilon_1; \beta_1) \tag{1}$$

The more objective measure  $H^0$ , which serves as a proxy for “true health”, is a vector of factors that affect true heterogeneity in health ( $X_2$ ), and a random error ( $\varepsilon_2$ ).

$$H^* = f_2(H^0, X_2, \varepsilon_2; \beta_2) \tag{2}$$

$H^S$  is an ordered categorical variable, and divergence in reporting styles may influence the group means of the index function and also change the cut-off points between the categories,  $j = 1, \dots, n$ .

$$H^S = j \leftrightarrow c_{j-1} < H^* \leq c_j, j = 1, \dots, n \tag{3}$$

The possible dependence of the cut-off points on  $X_j$  is given by the function  $g_j(\cdot)$ .

$$c_j = g_j(X_1; \beta_{1j}), j = 1, \dots, n - 1, c_0 = -\infty, c_n = \infty \tag{4}$$

Assuming a simple form for the approximation (3), the following equation can be thought of as reflecting true health as a measured component, which depends on  $H_0$ , and a further component, which depends on the “true” impacts of personal characteristics,  $X_2$ , on health, which is not captured by the more objective measure,  $H_0$ :

$$H^* = f(H^0; \alpha) + X_2' \beta_2 + \varepsilon_2 \tag{5}$$

Combining the above relationships results in the empirical equivalent of (3):

$$H^S = j \leftrightarrow g_{j-1}(X_1; \beta_{1j-1}) - X_2' \beta_2 < f(H^0; \alpha) + \varepsilon_2 \leq g_j(X_1; \beta_{1j}) - X_2' \beta_2 \tag{6}$$

Separate ordered response models are estimated by gender ( $k$ ), hence the above equation becomes:

$$H^S = j \leftrightarrow \delta_{j-1}^k < f(H^0; \alpha^k) + \varepsilon_2^k \leq \delta_j^k \tag{7}$$

This estimation equation takes into account differences in both index shifts, (represented by  $\alpha$ ), and cut-points (represented by  $\delta_j$ ) between the subgroups,  $k$ , of men and women.

To distinguish index shifts from cut-point shifts, the hypothesis that the  $\beta_{1j}$ ’s are identical for all categories (that is, for all values of  $j$ ) can be imposed. This is the null hypothesis, which implies that group-specific characteristics (such as age) do not influence perceptions about the location of cut-points between the categories of SRH. Hence, the rejection of the null hypothesis leads to the conclusion that the thresholds between SRH categories are affected by different reporting styles. In other words, reporting heterogeneity leads to cut-point shifts, which are more distorting than index shifts.

We estimate reporting heterogeneity in SRH by implementing Kerkhofs and Lindeboom's (1995) model through the use of the random effects generalised ordered probit (REGOPROB) procedure, developed by Pfarr, Schmid and Schneider (2011). This method was chosen because it takes advantage of the panel structure of the data and is the best available option for understanding patterns of reporting heterogeneity, given the range of health indicators in the dataset. The REGOPROB is an extension of the conventional probit regression method.

The equation for the conventional ordered probit estimation is given by:

$$\begin{aligned}
 Y_{it}^* &= X_{it}\beta + \varepsilon_{it} \\
 i &= 1, \dots, N \\
 t &= 1, \dots, T
 \end{aligned}
 \tag{8}$$

where  $Y_{it}^*$  is a latent dependent variable, which is assumed to be continuous (here, it represents true health);  $X_{it}$  is a vector of explanatory variables (here, more objective health measures and demographic variables);  $\beta$  is a vector of unknown parameters;  $\varepsilon_{it}$  is the error term;  $i$  is an index of individuals; and  $t$  is a time index.

As  $Y_{it}^*$  is unobserved, individuals provide a proxy,  $Y_{it}$  such that:

$$\begin{aligned}
 Y_{it} &= 0 \text{ if } Y_{it}^* \leq 0 \\
 &\dots \\
 Y_{it} &= J \text{ if } \tau_{J-1} \leq Y_{it}^*
 \end{aligned}
 \tag{9}$$

where  $Y_{it}$  is the categorical variable, SRH, with  $J$  ordered response options, and  $\tau$  is a vector of unknown cut-point parameters, to be estimated using the elements of the vector  $\beta$ .

The probit model generates a single index reflecting the underlying latent variable  $Y^*$ . Importantly, it assumes that a change in one of the explanatory variables shifts this index uniformly, where the slope coefficient  $\beta$  remains unchanged across individuals (or groups). This is referred to as the "parallel lines assumption". The ordered probit model therefore is constrained to reflect only index shifts. In contrast, the generalised ordered probit relaxes this assumption and allows for the cut-points between SRH categories to vary between individuals or groups of individuals, that is, for cut-point shifts. That is, it allows for variation in the slope coefficients  $\beta$ , as an alternative to index shifts.

The random effects ordered probit (REOP) takes the panel structure of the data into account by allowing for unobserved heterogeneity between individuals (Pfarr et al., 2011). The REOP assumes that individual-specific heterogeneity is distributed randomly across the population, and it provides cluster-specific coefficients. It can be expressed as follows:

$$Y_{it} = \mu + X_{it}\beta + \alpha_i + u_{it}
 \tag{10}$$

where  $\alpha_i$  is the individual-specific heterogeneity and  $u_{it}$  is an idiosyncratic error term.

The REGOPROB which we estimate utilises panel data to capture two kinds of heterogeneity: unobserved individual heterogeneity, which is assumed to be distributed randomly across the population; and observed heterogeneity, which is reflected in differences in the cut points and hence in the slope coefficients (Pfarr et al., 2011, 2012). The REGOPROB applies the iterative procedure designed by Pfarr et al.

(2011), which conducts a Wald test to test the parallel lines hypothesis for each explanatory variable. If this hypothesis is rejected, then different cut-points are estimated. As a final step, the REGOPROB procedure tests the hypothesis that the final model differs significantly from the REOP (in which the parallel lines assumption is taken to be valid) using a global Wald test (Pfarr et al., 2011).

In the Wald tests (as applied to individual variables or the overall model), when the null hypothesis of the absence of cut-point shifts is not rejected, then index shifts remain to be investigated. If there are differences in reported health between groups, this may be due to certain groups truly being healthier, or a result of reporting behaviour that displaces all categories of SRH in parallel, i.e. index shifts.

Kerkhofs and Lindeboom's (1995) model requires the inclusion of "objective" health measures (as in Eq. (6) above) against which self-reported measures can be benchmarked (Kerkhofs & Lindeboom, 1995; Pfarr et al., 2012).<sup>1</sup> The most objective measures would be provided by medical professionals, which is not feasible in large-scale surveys, and studies therefore mostly use self-reports of health conditions or symptoms. However, to the extent that self-reports contain measurement error, they are likely to be affected by heterogeneous reporting. We therefore refer to these benchmarks as "quasi-objective" health measures. These measures are typically constructed by aggregating a number of health conditions into one index, or by first estimating disability weights for individual conditions (e.g. Kerkhofs & Lindeboom, (1995); van Doorslaer & Jones, (2003); and Pfarr et al., (2012)).

The survey we analysed included a module where enumerators measured the respondent's blood pressure, height, weight, and waist circumference, but these "objective" indicators are compromised by extensive missing data and concerns about measurement error. Instead, and like earlier studies, we rely on "quasi-objective" health indicators. The survey did not ask the standard battery of questions that other studies have used to derive an overall health index or an overarching disability variable that could be used for estimating disability weights. Instead, and as we describe below, we include a range of indicators which capture health conditions, health behaviours, and healthcare utilisation to benchmark the adult's self-rated health.

## Data and Variables

The data for the study come from the first four waves<sup>2</sup> of the National Income Dynamics Study (NIDS), a large-sample nationally representative panel survey of South Africa (SALDRU, 2008; 2010; 2012; 2014). The sample is restricted to adults (aged 18 and over) who were resident household members in wave 1 of the panel, with separate sub-samples for men and women.

---

<sup>1</sup> Anchoring vignettes are likely to offer a more reliable method for comparing health perceptions (Bago d'Uva et al., 2008; Rossouw et al., 2018; Tubeuf, 2009). However, anchoring vignette questions are lengthy and time-consuming to administer, and hence they are not used in nationally representative surveys.

<sup>2</sup> The fifth wave of NIDS was not used because of substantial attrition, which necessitated refreshing the sample with additional respondents.



The main variable of interest is the individual's SRH, collected from the question: "How would you describe your health at present? Would you say it is excellent, very good, good, fair, or poor?" The two categories, "fair" and "poor", are combined to produce a four-level SRH measure because relatively few respondents throughout the panel chose the lowest category.

The health benchmarks for assessing SRH are captured through three sets of variables. The first is health conditions, which includes reporting on chronic conditions (whether individuals had "ever been told by a doctor, nurse or healthcare professional" that they have tuberculosis (TB), high blood pressure, diabetes, stroke, asthma, heart problems, and/or cancer). We divide individuals into three categories: those with no known chronic conditions, those with exactly one, and those with two or more. We also model information from a follow-up question, which probed whether the respondent had another major illness or disability (e.g. sight, hearing, or speech, psychological or psychiatric disorders, HIV/AIDS, epilepsy, and Alzheimer's disease). We generate a binary indicator that equals one if the individual reported one or more other condition or disability. In addition, a mental health indicator is estimated using responses to ten questions, derived from the depression scale developed by the Centre for Epidemiological Studies (CES-D10) (Radloff, 1977). The scale has been validated for use with Zulu, Xhosa, and Afrikaans-speaking individuals in South Africa (Baron et al., 2017). We create a binary indicator equal to one if the CES-D10 score was 10 or higher (from a maximum score of 30). To capture health conditions, we also include information collected from a question about short-term health, where respondents are asked if, in the last 30 days, they had experienced any of 24 symptoms listed, ranging from a headache to a serious injury. Responses are combined into a single binary variable, showing whether the individual had experienced at least one of the symptoms.

The second quasi-objective health indicators are the adult's health behaviours, which are derived from questions on alcohol consumption, smoking, and physical activity. For alcohol consumption, adults were asked "How often do you drink alcohol?", and we group the responses into a binary indicator equal to one if the individual consumed alcohol three or more times per week. This classification is based roughly on the "hazardous drinking score", a screening tool developed by the United States' Substance Abuse and Mental Health Services Administration (Center for Behavioral Health Statistics and Quality, 2018)<sup>3</sup>. Information on smoking is a binary variable, reflecting whether the individual currently smokes cigarettes. Physical exercise is also gauged via a single question, "How regularly do you exercise?" The WHO recommends that adults do at least 150 min of moderate aerobic physical activity throughout the course of a week (WHO, 2010). Since only 27% of men and 12% of women from our sample met this criterion, we apply a less stringent measure of whether the individual exercises at least once a week.

The final quasi-objective health measure comes from information on healthcare utilisation, and specifically a question about the time elapsed since the respondent's

<sup>3</sup> Drinking three or more times per week need not be hazardous for health; this depends on the quantities consumed (Center for Behavioral Health Statistics and Quality, 2018).

last medical consultation. As some of the eight possible time intervals in the NIDS questionnaire (from “never” to “in the last 30 days”) were rarely chosen, the indicator employed here was reduced to three categories, namely: “five or more years ago or never”, “one to four years ago”, and “in the last year”.

The NIDS instrument also captured some measured indicators, which would provide more “objective” benchmarks of the adult’s health. Specifically, enumerators were instructed to measure each respondent’s weight, height, and waist circumference, and to take two measures of blood pressure (Ardington & Case, 2009; Ardington & Gasealahwe, 2012). However, the use of these indicators is compromised by missing data and concerns with measurement error. We therefore assess reporting heterogeneity first excluding and then including two “objective” indicators: the body mass index (BMI) category and hypertension.

The body mass index is a widely used measure of nutritional status and is calculated as the individual’s weight (in kilograms) divided by the square of her height (in metres). We follow the common practice of bracketing this measure into four categories, according to the WHO classification: underweight (BMI below 18.5), normal (BMI between 18.5 and 24.9), overweight (BMI between 25 and 29.9), and obese (BMI above 30). These categories are defined only for ages 20 and older (WHO, 2000). For respondents with two non-missing measurements, the average weight and height are used for the BMI calculation. Observations with a BMI below 10 or above 50 in both assessments are omitted from the analysis, as these likely reflect measurement errors. The use of the BMI indicator warrants further caution because it is subject to non-random patterns of missing data. For example, in wave 1, white women were three times more likely to have missing weight and height data, compared to African women, while the proportion of missing values for white men was twice that for African men (Ardington & Case, 2009). In the pooled sample of waves 1 to 4, the share of white women for whom the BMI cannot not be calculated (mostly because of missing values) is 17%, compared to 12% of African women; the corresponding percentages for men are approximately the same.

The indicator for hypertension is calculated by averaging the two blood pressure measurements and then creating a binary indicator for severe hypertension if the average systolic reading was above 179 or the average diastolic reading was above 109. Similar to the BMI indicator, the pattern of missing blood pressure measurements is non-random. In the pooled sample of waves 1 to 4, the percentage of missing blood pressure readings for white women is 20%, compared to 9% for African women; for men, the corresponding percentages are 19% and 8%.

The objective of the regression analysis is to test whether reporting heterogeneity exists between adults with different demographic and socioeconomic characteristics, distinguishing the analysis by gender. These characteristics encompass a range of individual-level, geographical location and household-level descriptors. The individual characteristics include: a quadratic in age (as the path of individuals’ health perceptions over time is assumed to be non-linear); three of the four population groups commonly delineated in official South African surveys (African, coloured and white, with Asian/Indian respondents excluded because of small sub-sample sizes); marital status (“single” or never married, “married or cohabiting”, and “divorced, separated, or widowed”); education, which is included as a four-level

variable with categories based on years of schooling (no schooling or incomplete primary (up to 6 years of education), primary or incomplete secondary (between 7 and 11 years), secondary or incomplete tertiary (between 12 and 14 years), and some level of tertiary education (15 years or more).

Given the persistence of apartheid geographies, we consider the geographical type (geo-type) of residence, distinguishing between: urban areas (defined as built-up cities and towns, including formal and informal settlements); traditional areas (villages on community-owned land, which are under traditional leaders' authority and were part of the former "homelands" or "Bantustans"); and farms (representing commercial farming areas) (Chinhema et al., 2016). Finally, the regressions include a measure of real monthly household expenditure per capita, which is adjusted for inflation (according to the price level prevailing in November 2014, the last month of interviews for wave 4).

## Descriptive Statistics

The descriptive statistics for the health variables used in the analysis are presented in Table 1, while the demographic and socio-economic characteristics of men and women are shown in Table 2. As expected, more women than men rate their health as fair or poor, while the other self-reported health indicators describe significantly higher morbidity among women than men: women are more likely to have been diagnosed with at least one chronic condition, to have a disability, to have experienced symptoms of illness recently, or to suffer from (at least mild) depression.

Substantial gender differences in lifestyle behaviours are reflected both in the higher rates of smoking and alcohol consumption among men, and in men's better exercise habits. Health-care seeking behaviour also has the expected gender pattern, where women are more likely than men to have consulted a health practitioner within the last four years.

Table 2 shows that women are older than men on average, being never married is the modal marital status, and there are no gender differences in tertiary educational attainment although a small gender gap remains among the least educated (who tend to be older than the rest of the sample).

Consistent with findings from other research, women are significantly more likely than men to live in households with lower expenditure per capita (Posel & Hall, 2021), and women remain more likely than men to live in traditional rural areas. Differences in the residential location of women and men reflect historical, and continuing, patterns of predominantly male labour migration from rural to urban areas in South Africa (Posel, 2010).

## Results

Tables 3 and 4 present the results from random effects generalised ordered probit (REGOPROB) regressions for men and women respectively. As explained earlier, the estimations use the quasi-objective health indicators, as well as variables

**Table 1** Health indicators: descriptions and summary statistics

Variable	Definition		Men	Women
Self-rated health (4-level)	Self-rated health 1 if fair or poor 2 if good 3 if very good 4 if excellent	Fair or poor	0.104 (0.004)	0.160*** (0.005)
		Good	0.234 (0.006)	0.262*** (0.005)
		Very good	0.292 (0.007)	0.277** (0.006)
		Excellent	0.371 (0.008)	0.301*** (0.007)
		Chronic conditions	Number of diagnosed chronic conditions	None <sup>#</sup>
		One	0.047 (0.003)	0.059*** (0.003)
		Two or more	0.099 (0.005)	0.180*** (0.006)
Other condition	Other major illnesses/disability	No <sup>#</sup>	0.949 (0.003)	0.934*** (0.003)
		Yes	0.051 (0.003)	0.066*** (0.003)
Regular drinker	Drinks alcohol $\geq 3$ times/week	No <sup>#</sup>	0.938 (0.003)	0.987*** (0.002)
		Yes	0.062 (0.003)	0.013*** (0.002)
Smoker	Smokes currently	No <sup>#</sup>	0.635 (0.010)	0.912*** (0.009)
		Yes	0.365 (0.010)	0.088*** (0.009)
Exercise	Frequency of physical exercise	Never/ infrequently <sup>#</sup>	0.659 (0.008)	0.832*** (0.009)
		$\geq 1$ /week	0.341 (0.008)	0.168*** (0.009)
Depressed	CES-D10 $\geq 10$	No <sup>#</sup>	0.763 (0.007)	0.715*** (0.008)
		Yes	0.237 (0.007)	0.285*** (0.008)
Last medical consultation	Years since last medical consultation	$\geq 5$ years or never <sup>#</sup>	0.217 (0.010)	0.135*** (0.006)
		1 to 4 years ago	0.424 (0.010)	0.342*** (0.008)
		Last year	0.360 (0.009)	0.523*** (0.008)

**Table 1** (continued)

Variable	Definition		Men	Women
Symptoms	Number experienced in past 30 days	None <sup>#</sup>	0.494 (0.007)	0.388*** (0.007)
		≥ 1	0.506 (0.007)	0.612*** (0.007)
BMI category	(≥ 20 years only) Underweight: 10.00—18.49 Normal: 18.50—24.99 Overweight: 25.00—29.99 Obese: ≥ 30.00	Underweight	0.079 (0.005)	0.029*** (0.002)
		Normal <sup>#</sup>	0.535 (0.013)	0.288*** (0.007)
		Overweight	0.242 (0.010)	0.272*** (0.006)
		Obese	0.143 (0.008)	0.410*** (0.010)
Hypertension	Severe hypertension (systolic blood pressure > 179; diastolic blood pressure > 109)	No <sup>#</sup>	0.992 (0.001)	0.988*** (0.001)
		Yes	0.008 (0.001)	0.012*** (0.001)
N			20,489	30,583

Source: Own calculations based on wave 1 to 4 of NIDS

\*\*\* Mean or proportion difference between men and women is statistically significant at the 1% level ( $p < 0.01$ ); \*\* 5% level ( $p < 0.05$ ). # denotes the omitted category for categorical variables. Robust standard errors are shown in parentheses. Estimates are weighted using the post-stratified weights

reflecting lifestyle and health-care seeking behaviour, as the more objective health benchmarks against which reporting styles can be assessed. An alternative specification of the regressions, which includes the two measured health indicators, is reported in [Appendix 1](#).

For each gender, the REGORPOB estimates three equations representing binary choices between SRH categories. The first equation represents the comparison between the worst health category ("fair or poor") and the three remaining options ("good", "very good" and "excellent"). The second equation reflects the choice between the bottom two categories ("good" and "fair or poor") and the top two ("excellent" and "very good"). The last equation juxtaposes the best category ("excellent") against the others.

Cut-point shifts are identified when an individual characteristic has different and statistically significant coefficients in one or more of the REGOPROB equations. Such a finding implies that the cut points between the SRH categories are not equally spaced for individuals with the corresponding characteristic, compared to the reference category (or for a marginal change if the characteristic is a continuous variable) (Schneider et al., 2012). For example, the same degree of deterioration, say, in the underlying "true health" index will lead an individual from one group to switch to a lower SRH category, while an individual from another group will leave their health rating unchanged. An index shift

**Table 2** Demographic and socioeconomic variables

Variable	Definition		Men	Women
Age	Age at last birthday		37.996 (0.434)	39.906*** (0.314)
Race	Population group	African <sup>#</sup>	0.807 (0.024)	0.792* (0.024)
		Coloured	0.088 (0.018)	0.095* (0.018)
		White	0.105 (0.017)	0.112 (0.018)
Marital status		Never married <sup>#</sup>	0.541 (0.016)	0.459*** (0.012)
		Married/cohabiting	0.399 (0.014)	0.390 (0.011)
		Divorced/separated/widowed	0.059 (0.007)	0.151*** (0.005)
Education	None or incomplete primary (0–6 yrs) Primary or incomplete secondary (7–11 yrs) Secondary or incomplete tertiary (12–14 yrs) Tertiary (15 years or more)	None <sup>#</sup>	0.183 (0.010)	0.204** (0.010)
		Primary	0.475 (0.012)	0.463 (0.010)
		Secondary	0.303 (0.013)	0.292 (0.010)
		Tertiary	0.039 (0.009)	0.040 (0.007)
Expenditure	Real household expenditure per capita quintiles	Quintile 1 <sup>#</sup>	0.097 (0.007)	0.145*** (0.008)
		2	0.105 (0.007)	0.152*** (0.007)
		3	0.132 (0.006)	0.168*** (0.006)
		4	0.217 (0.009)	0.211 (0.008)
		5	0.448 (0.019)	0.324*** (0.019)
Geo-type	Type of geographical area	Urban <sup>#</sup>	0.672 (0.027)	0.633*** (0.028)
		Traditional	0.279 (0.026)	0.331*** (0.028)
		Farms	0.049 (0.008)	0.036*** (0.006)
N			20,489	30,583

Source: Own calculations based on Wave 1 to Wave 4 of NIDS

\*\*\* Mean or proportion difference between men and women is statistically significant at the 1% level ( $p < 0.01$ ), \*\* 5% level ( $p < 0.05$ ), or \* 10% level ( $p < 0.1$ ). # denotes the omitted category for categorical variables. Robust standard errors are shown in parentheses. Estimates are weighted using the post-stratified weights

**Table 3** Reporting heterogeneity in self-rated health: Men

	Dependent variable: Four-level self-rated health					
	1 vs. 2–4		1–2 vs. 3–4		1–3 vs. 4	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
Age	-0.029***	(0.005)	-0.010***	(0.004)	-0.008**	(0.004)
Age <sup>2</sup>	0.007	(0.005)	-0.009**	(0.004)	-0.010**	(0.004)
Race						
Coloured	0.107***	(0.042)	0.080**	(0.032)	-0.029	(0.032)
White	0.369***	(0.072)	0.138**	(0.055)	-0.075	(0.056)
Education						
Primary	0.196***	(0.025)	0.196***	(0.025)	0.196***	(0.025)
Secondary	0.309***	(0.030)	0.309***	(0.030)	0.309***	(0.030)
Tertiary	0.448***	(0.066)	0.448***	(0.066)	0.448***	(0.066)
Expenditure quintiles						
2	0.059*	(0.030)	0.059*	(0.030)	0.059*	(0.030)
3	0.140***	(0.030)	0.140***	(0.030)	0.140***	(0.030)
4	0.118***	(0.029)	0.118***	(0.029)	0.118***	(0.029)
5	0.180***	(0.030)	0.180***	(0.030)	0.180***	(0.030)
Marital status						
Married/cohabiting	0.114***	(0.025)	0.114***	(0.025)	0.114***	(0.025)
Divorced/Separated/Widowed	0.052	(0.045)	0.052	(0.045)	0.052	(0.045)
Geo-type						
Traditional	0.127***	(0.033)	0.027	(0.025)	-0.080***	(0.025)
Farms	0.051	(0.032)	0.051	(0.032)	0.051	(0.032)
Chronic conditions						
1	-0.282***	(0.037)	-0.282***	(0.037)	-0.282***	(0.037)
≥ 2	-0.460***	(0.030)	-0.460***	(0.030)	-0.460***	(0.030)
Other condition/disability	-0.733***	(0.043)	-0.679***	(0.045)	-0.538***	(0.057)
Regular drinker	-0.065*	(0.035)	-0.065*	(0.035)	-0.065*	(0.035)
Smoker	-0.092***	(0.019)	-0.092***	(0.019)	-0.092***	(0.019)
Exercise	0.102***	(0.019)	0.102***	(0.019)	0.102***	(0.019)
Depressed	-0.281***	(0.030)	-0.247***	(0.023)	-0.157***	(0.024)
Symptoms	-0.628***	(0.032)	-0.394***	(0.021)	-0.368***	(0.021)
Last medical consultation						
1–4 years	-0.090*	(0.047)	-0.081***	(0.027)	-0.199***	(0.025)
Last year	-0.670***	(0.045)	-0.472***	(0.030)	-0.405***	(0.029)
Wave						
2	0.288***	(0.039)	0.211***	(0.030)	0.138***	(0.029)
3	0.341***	(0.038)	0.026	(0.028)	-0.017	(0.029)
4	0.366***	(0.037)	0.031	(0.027)	-0.020	(0.028)
Constant	2.706***	(0.119)	1.042***	(0.085)	0.193**	(0.085)
N	20,489					
Rho	0.079*** (0.010)					

Source: Own calculations based on Wave 1 to 4 of NIDS

The sample is men aged 18 and older. \*\*\* Significant at the 1% level ( $p < 0.01$ ), \*\* 5% level ( $p < 0.05$ ), \* 10% level ( $p < 0.1$ )

**Table 4** Reporting heterogeneity in self-rated health: Women

	Dependent variable: Four-level self-rated health					
	1 vs. 2–4		1–2 vs. 3–4		1–3 vs. 4	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
Age	-0.030***	(0.003)	-0.013***	(0.003)	-0.015***	(0.003)
Age <sup>2</sup>	0.009***	(0.003)	-0.006**	(0.003)	-0.002	(0.003)
Race						
Coloured	0.154***	(0.033)	0.168***	(0.027)	0.025	(0.029)
White	0.454***	(0.062)	0.327***	(0.049)	0.063	(0.050)
Education						
Primary	0.200***	(0.020)	0.200***	(0.020)	0.200***	(0.020)
Secondary	0.360***	(0.025)	0.360***	(0.025)	0.360***	(0.025)
Tertiary	0.448***	(0.053)	0.448***	(0.053)	0.448***	(0.053)
Expenditure quintiles						
2	0.027	(0.022)	0.027	(0.022)	0.027	(0.022)
3	-0.017	(0.030)	-0.000	(0.025)	0.086***	(0.027)
4	0.008	(0.030)	-0.002	(0.026)	0.100***	(0.027)
5	0.029	(0.034)	0.059**	(0.029)	0.195***	(0.029)
Marital status						
Married/cohabiting	0.064***	(0.018)	0.064***	(0.018)	0.064***	(0.018)
Divorced/Separated/Widowed	0.035	(0.024)	0.035	(0.024)	0.035	(0.024)
Geo-type						
Traditional	0.097***	(0.024)	0.082***	(0.020)	-0.029	(0.021)
Farms	0.000	(0.043)	0.069**	(0.035)	-0.065*	(0.038)
Chronic conditions						
1	-0.253***	(0.028)	-0.253***	(0.028)	-0.253***	(0.028)
≥2	-0.399***	(0.020)	-0.399***	(0.020)	-0.399***	(0.020)
Other condition/disability	-0.696***	(0.032)	-0.502***	(0.034)	-0.373***	(0.042)
Regular drinker	-0.128**	(0.059)	-0.128**	(0.059)	-0.128**	(0.059)
Smoker	-0.081***	(0.027)	-0.081***	(0.027)	-0.081***	(0.027)
Exercise	0.059***	(0.020)	0.059***	(0.020)	0.059***	(0.020)
Depressed	-0.300***	(0.021)	-0.221***	(0.018)	-0.131***	(0.020)
Symptoms	-0.613***	(0.025)	-0.416***	(0.017)	-0.440***	(0.018)
Last medical consultation						
1–4 years	0.042	(0.040)	-0.084***	(0.026)	-0.161***	(0.026)
Last year	-0.509***	(0.037)	-0.442***	(0.026)	-0.363***	(0.026)
Wave						
2	0.234***	(0.028)	0.270***	(0.023)	0.191***	(0.025)
3	0.365***	(0.027)	0.133***	(0.023)	0.035	(0.024)
4	0.498***	(0.027)	0.207***	(0.022)	0.063***	(0.024)
Constant	2.638***	(0.090)	1.036***	(0.068)	0.197***	(0.070)
N	30,583					
Rho	0.058*** (0.007)					

Source: Own calculations based on wave 1 to 4 of NIDS

The sample is women aged 18 and older. \*\*\* Significant at the 1% level ( $p < 0.01$ ), \*\* 5% level ( $p < 0.05$ ), \* 10% level ( $p < 0.1$ )



occurs where a characteristic leads to statistically significant coefficients that are the same in all three of the equations (Schneider et al., 2012). The Wald test for each overall model tests the validity of the parallel lines assumption for all variables in the model (Pfarr et al., 2011).

Among men, education, household expenditure, and marital status cause index shifts. Education is positively related to SRH. More household resources also have a favourable effect on health, which is reflected in the positive index shifts for the second to fifth expenditure quintiles of household expenditure, compared to the bottom quintile. However, the distinction between the bottom two quintiles is statistically significant only at the 10% level. Married men rate their health more positively than those who are unattached, while there is no significant difference between the perceptions of never married and formerly married men.

For men, cut-point shifts are associated with age, race and location. As expected, the coefficients on the two age variables reflect a negative effect. The relationship is stronger and linear in the first equation (as older men are more likely to rate their health as less-than-good), but it weakens and accelerates with age when the lower SRH categories are considered. The positive and strongly significant coefficients on the race variables across the first two equations indicate that coloured and white men tend to report being in better health than the reference category, African men, with larger differences for white men. The cut-point shift is biggest in the first equation, indicating that compared to African men, it would take a larger health deterioration for coloured men to choose a less-than-good rating, with this pattern amplified for white men. Finally, men in traditional rural areas are significantly more likely than urban dwellers to report better than less-than-good health, but less likely to report excellent health.

For women, the variables representing parallel shifts for all SRH categories are the same as for men, except for the expenditure quintile dummy variables. The health benefits of education are similar in magnitude for both genders, while marriage has a much smaller, but still significant, positive effect on women's health perceptions.

The drivers of reporting heterogeneity for women are age, race, household expenditure and location. In contrast to men, a non-linear pattern in age is evident in the first equation, which shows that women are more likely to perceive their health as "fair or poor" with age, but this effect slows over time. This is consistent with findings that women's healthcare costs in South Africa are lower than men's from about age 40 (McLeod, 2012). McLeod offers no explanation, but possible mechanisms include lower maternal healthcare costs as women age, while older men are more likely than women to suffer from life-threatening conditions. However, as is the case with men, the likelihood of women reporting worse health increases over time when the comparison is between poor or fair health, and good or excellent health. The coefficients on the race dummy variables describe a similar pattern for women as for men, although the significant associations are stronger for women. Like men, women in traditional areas are also more optimistic about their health, and even more likely to report very good or excellent health (although they also avoid reporting their health as excellent).

The most notable gender difference in reporting behaviours concerns the association between SRH and household expenditure. While higher household expenditure implies uniform shifts into better health categories for men, it has a weaker and

heterogeneous effect for women. The coefficients for expenditure quintiles are largely insignificant in the first two equations, but are highly significant in the third, indicating that more resources increase women's reporting of excellent health in particular.

Considering the overall models for both men and women, the Wald tests fail to reject the null hypothesis that the parallel lines assumption is valid, at the 1% and 5% significance levels, but not at the 10% level. The p-values for men and women are 0.059 and 0.077 respectively.

The results from the regressions which include measures for BMI and hypertension show similar patterns of reporting heterogeneity ([Appendix 1](#)). For women, the Wald test also fails to reject the null hypothesis that the parallel lines assumption is valid, but again only at the 1% and 5% significance levels (with a p-value of 0.06). For men, the hypothesis cannot be rejected at all conventional significance levels (with a p-value of 0.19). Women's reporting behaviour therefore reflects more variability and heterogeneity, which is consistent with the international literature ([Idler, 2003](#)). The inclusion of the measured indicators reduces the goodness of fit of the overall model somewhat (with lower values for rho), which may signal inconsistencies in health reporting over time and discrepancies between self-reported and measured indicators. For example, using data from Wave 1 of NIDS, Ardington and Case ([2009](#)) show that for those individuals whose blood pressure measurements were in the range indicating mild hypertension, only 49% of women and 26% of men reported ever being diagnosed with the condition by a health professional.

The quasi-objective and measured indicators are included as controls, and therefore they cannot be interpreted in the light of reporting heterogeneity. However, there are some noteworthy gender differences in their relationship to SRH. The more objective benchmark indicators have similar associations to SRH for both men and women, apart from chronic conditions, which appear to take a greater toll on men's health, and regular exercise benefits men more. Men and women who are measured as being underweight both report significantly lower SRH, while being overweight has a small but significant positive effect on SRH. A BMI in the 'obese' range appears to be protective of poor health, but this is only the case for women. This result is consistent with the positive association between socioeconomic indicators and obesity reported by Ardington and Gasealahwe ([2012](#)), and it may stem partly from cultural perceptions among Africans that by being more 'curvaceous', women signal that they are materially and physically better-off ([Wittenberg, 2013](#)).

## Tests for Attrition Bias

As the study analysed panel data, we used the procedure designed by Beckett, Gould, Lillard, and Welch (known as the BGLW test) to test for possible attrition bias in the reporting heterogeneity models ([Beckett et al., 1988](#)). The test investigates the effects of attrition in the next period on the outcome variable in the present period. The null hypothesis of no attrition bias is equivalent to testing whether the binary attrition variable and its interaction terms with the regressors are not jointly statistically significant.

The BGLW test was conducted for each pair of consecutive waves, as non-random patterns of (positive) attrition were recorded between waves 1 and 2, and also between waves 3 and 4 (Brown et al., 2012; Chinhema et al., 2016). In addition, the test was applied in the conventional way and in reverse for Waves 2 and 3 because of the negative attrition recorded between wave 2 and wave 3 of NIDS (De Villiers et al., 2013). The 'reverse' test for the effect of the negative attrition between wave 2 and wave 3 was conducted by setting wave 3 as the 'baseline wave' and including a binary variable for individuals who were interviewed in wave 3 but not in wave 2, as well as interactions of this variable with the other explanatory variables, into the regression results for wave 3.

The BGLW test results are presented in Table 5. They suggest that attrition had only a limited effect on the regression results: the null hypothesis of no attrition bias is supported, except for wave 1 to wave 2 attrition among men (at all conventional significance levels) and wave 2 to wave 3 attrition among women (at the 5% and 10% significance levels, but not at the 1% level). (The complete results of the BGLW test are available from the authors.)

### Conclusion

Health is a core constituent of an individual's quality of life, and a key objective of empirical work on health is to document and investigate inequalities that differentially limit people's ability to participate meaningfully in society. However, its multifaceted nature makes it difficult to measure health at the level of the population.

Self-rated health (SRH) is "perhaps the most frequently used measure of health in the social sciences" (Denton et al., 2004: 2597) – it is often included in surveys that collect a wide range of other information on individuals and the households in which they live; and as an individual assessment, SRH arguably comes closest, both conceptually and practically, to the broad definition of health used by the WHO (1948). Nevertheless, the subjective nature of the measure also undermines its credibility as a reliable indicator that can be compared across groups of individuals who may have different reporting styles, or different understandings of what constitutes good and bad health.

This study adds to a small but growing literature that investigates the reliability of SRH in a developing country context. In contrast to many other studies, we tested for reporting heterogeneity in SRH through an analysis of longitudinal data for South Africa (collected in the National Income Dynamics Study), where we

**Table 5** BGLW test for attrition in regression models

	Men			Women		
	Chi2	df	p-value	Chi2	df	p-value
Wave 1–2	40.840	22.000	0.009	28.329	22.000	0.165
Wave 2–3	21.117	22.000	0.514	35.155	22.000	0.037
Wave 3–2	27.107	22.000	0.207	27.029	22.000	0.210
Wave 3–4	21.937	22.000	0.464	26.068	22.000	0.249

Source: Own calculations using NIDS Waves 1 to 4

implemented Kerkhofs and Lindeboom's (1995) model using the random effects generalised ordered probit procedure subsequently developed by Pfarr et al. (2011).

An important advantage of the method we used is that it exposes differences in reporting behaviour in the form of both index and cut-point shifts, where an index shift reflects a parallel shift in the whole distribution of SRH responses, while the more distorting cut-point shift implies unequal spacing between response options, from the worst to the best. The use of longitudinal data allowed us to control for any unobserved individual heterogeneity, which could bias the analysis of reporting behaviour, and the comprehensive nature of the survey permitted the consideration of a wide array of observable factors that could influence health perceptions.

Our analysis focussed on gender differences in reporting on SRH, to assess whether patterns of reporting heterogeneity differed systematically between women and men, thereby compromising gender comparisons of morbidity that are based on subjective assessments. As is regularly documented across the world, women in South Africa provide significantly lower SRH reports than men. The estimations in this study identified that women's reporting behaviour on SRH exhibited more variability and heterogeneity than that of men. This has also been documented in the international literature (Idler, 2003) and would be consistent with the argument that social factors have a stronger influence on women's health perceptions than on men's health perceptions.

However, the analysis did not find evidence that SRH overall, for either women or men, was compromised by distorting cut-point shifts in response options. Although the estimations identified several gender differences in reporting styles, these were mostly in the form of index shifts. The findings from this study therefore corroborate the reliability of SRH, even in the diverse South African context, especially in research based on panel data. Nonetheless, they also affirm the importance of testing for reporting heterogeneity by gender, particularly when gender differences in health are the focus of the research.

Notwithstanding the advantages of the method and the data analysed, there were also limitations to the study that must be acknowledged. First, to benchmark SRH when testing for reporting heterogeneity, models typically include a more objective health measure. The survey we analysed collected information to derive two indicators (the BMI and hypertension) which would provide objective benchmarks, but these indicators are compromised by measurement error and non-random patterns of missing data. Like other studies, we therefore relied on "quasi-objective" indicators based on individual self-reports to a series of questions. However, in contrast to other studies, we were not able to aggregate this information into a single index; rather to benchmark SRH, we included a range of covariates which captured health conditions, health behaviours, and healthcare utilisation. Second, attrition affects the longitudinal data we analysed, although there was evidence of attrition bias in only one of the wave transitions, for both women and men. In the South African context, future research could also assess the reliability of SRH for drawing comparisons between "population group" or "race", to further probe whether sustained inequality in access to economic and health resources influences the formation of health perceptions.

**Appendix 1**

**Table 6** Reporting heterogeneity in SRH: Men (including measured indicators)

	Dependent variable: Four-level self-rated health					
	1 vs. 2-4		1-2 vs. 3-4		1-3 vs. 4	
	Coeff	SE	Coeff	SE	Coeff	SE
Age	-0.032***	(0.006)	-0.013***	(0.004)	-0.012**	(0.005)
Age <sup>2</sup>	0.010*	(0.006)	-0.005	(0.005)	-0.005	(0.005)
Race						
Coloured	0.108**	(0.046)	0.079**	(0.035)	-0.011	(0.036)
White	0.350***	(0.083)	0.109*	(0.062)	-0.124*	(0.064)
Education						
Primary	0.165***	(0.027)	0.165***	(0.027)	0.165***	(0.027)
Secondary	0.287***	(0.033)	0.287***	(0.033)	0.287***	(0.033)
Tertiary	0.365***	(0.072)	0.365***	(0.072)	0.365***	(0.072)
Expenditure quintiles						
2	0.094***	(0.034)	0.094***	(0.034)	0.094***	(0.034)
3	0.182***	(0.034)	0.182***	(0.034)	0.182***	(0.034)
4	0.166***	(0.033)	0.166***	(0.033)	0.166***	(0.033)
5	0.212***	(0.034)	0.212***	(0.034)	0.212***	(0.034)
Marital status						
Married/cohabiting	0.094***	(0.026)	0.094***	(0.026)	0.094***	(0.026)
Divorced/Separated/Widowed	0.028	(0.049)	0.028	(0.049)	0.028	(0.049)

Table 6 (continued)

	Dependent variable: Four-level self-rated health					
	1 vs. 2–4		1–2 vs. 3–4		1–3 vs. 4	
	Coeff	SE	Coeff	SE	Coeff	SE
Geo-type						
Traditional Farms	0.130*** (0.036)		0.026 (0.027)		-0.064** (0.027)	
	0.059* (0.035)		0.059* (0.035)		0.059* (0.035)	
Chronic conditions						
1	-0.278*** (0.041)		-0.278*** (0.041)		-0.278*** (0.041)	
≥ 2	-0.475*** (0.034)		-0.475*** (0.034)		-0.475*** (0.034)	
Other condition/disability	-0.702*** (0.049)		-0.632*** (0.050)		-0.499*** (0.063)	
Regular drinker	-0.075** (0.038)		-0.075** (0.038)		-0.075** (0.038)	
Smoker	-0.076*** (0.021)		-0.076*** (0.021)		-0.076*** (0.021)	
Exercise	0.097*** (0.021)		0.097*** (0.021)		0.097*** (0.021)	
Depressed	-0.262*** (0.033)		-0.234*** (0.026)		-0.149*** (0.027)	
Symptoms	-0.592*** (0.036)		-0.390*** (0.023)		-0.370*** (0.023)	
Last medical consultation						
1–4 years	-0.088* (0.053)		-0.089*** (0.053)		-0.210*** (0.028)	
Last year	-0.673*** (0.050)		-0.467*** (0.050)		-0.428*** (0.032)	
BMI category						
Underweight	-0.199*** (0.033)		-0.199*** (0.033)		-0.199*** (0.033)	
Overweight	0.132*** (0.039)		0.051* (0.029)		-0.007 (0.029)	
Obese	-0.012 (0.032)		-0.012 (0.032)		-0.012 (0.032)	
Hypertension	0.011 (0.088)		0.011 (0.088)		0.011 (0.088)	

Table 6 (continued)

		Dependent variable: Four-level self-rated health					
		1 vs. 2-4		1-2 vs. 3-4		1-3 vs. 4	
Wave		Coeff	SE	Coeff	SE	Coeff	SE
2		0.272***	(0.045)	0.206***	(0.045)	0.106***	(0.034)
3		0.327***	(0.042)	0.043	(0.042)	-0.017	(0.032)
4		0.362***	(0.040)	0.048	(0.040)	-0.015	(0.031)
Constant		2.758***	(0.144)	1.095***	(0.144)	0.272***	(0.105)
N		16,290					
Rho		0.072***	(0.011)				

Source: Own calculations based on wave 1 to 4 of NIDS

The *p*-value from the Wald test of the parallel lines assumption for the final model is 0.191, hence the null hypothesis that the final model does not violate the parallel lines assumption cannot be rejected at all conventional significance levels

**Table 7** Reporting heterogeneity in SRH: Women (including measured indicators)

Dependent variable: Four-level self-rated health						
1 vs. 2-4		1-2 vs. 3-4		1-3 vs. 4		
Coeff	SE	Coeff	SE	Coeff	SE	
Age	-0.038***	(0.004)	-0.016***	(0.003)	-0.013***	(0.004)
Age <sup>2</sup>	0.016***	(0.004)	-0.004	(0.003)	-0.004	(0.004)
Race						
Coloured	0.170***	(0.037)	0.144***	(0.031)	0.023	(0.032)
White	0.432***	(0.072)	0.322***	(0.056)	0.076	(0.058)
Educ						
Primary	0.174***	(0.021)	0.174***	(0.021)	0.174***	(0.021)
Secondary	0.325***	(0.027)	0.325***	(0.027)	0.325***	(0.027)
Tertiary	0.396***	(0.058)	0.396***	(0.058)	0.396***	(0.058)
Expenditure quintiles						
2	0.041 *	(0.024)	0.041 *	(0.024)	0.041 *	(0.024)
3	-0.028	(0.033)	0.008	(0.028)	0.101***	(0.030)
4	0.009	(0.034)	0.006	(0.029)	0.129***	(0.030)
5	0.007	(0.038)	0.079**	(0.032)	0.229***	(0.033)
Marital status						
Married/cohabiting	0.046**	(0.019)	0.046**	(0.019)	0.046**	(0.019)
Divorced/Separated/ Widowed	0.029	(0.026)	0.029	(0.026)	0.029	(0.026)
Geo-type						
Traditional	0.066**	(0.027)	0.055**	(0.022)	-0.024	(0.023)
Farms	-0.025	(0.047)	0.066*	(0.038)	-0.058	(0.042)



Table 7 (continued)

Dependent variable: Four-level self-rated health						
	1 vs. 2-4		1-2 vs. 3-4		1-3 vs. 4	
	Coeff	SE	Coeff	SE	Coeff	SE
<b>Chronic conditions</b>						
1	-0.258***	(0.031)	-0.258***	(0.031)	-0.258***	(0.031)
≥ 2	-0.403***	(0.022)	-0.403***	(0.022)	-0.403***	(0.022)
Other condition/disability	-0.685***	(0.035)	-0.506***	(0.037)	-0.401***	(0.046)
Regular drinker	-0.113*	(0.065)	-0.113*	(0.065)	-0.113*	(0.065)
Smoker	-0.076**	(0.030)	-0.076**	(0.030)	-0.076**	(0.030)
Exercise	0.045*	(0.023)	0.045*	(0.023)	0.045*	(0.023)
Depressed	-0.295***	(0.023)	-0.231***	(0.020)	-0.141***	(0.022)
Symptoms	-0.626***	(0.028)	-0.410***	(0.019)	-0.431***	(0.020)
<b>Last medical consultation</b>						
1-4 years		-0.009		-0.102***		-0.184***
Last year		-0.564***		-0.451***		-0.388***
<b>BMI category</b>						
Underweight		-0.161***	(0.041)	-0.161***	(0.041)	-0.161***
Overweight		0.140***	(0.030)	0.019	(0.024)	-0.036
Obese		0.123***	(0.028)	0.019	(0.023)	-0.010
Hypertension		-0.123**	(0.059)	-0.123**	(0.059)	-0.123**
<b>Wave</b>						
2		0.254***	(0.032)	0.290***	(0.027)	0.192***
3		0.390***	(0.030)	0.159***	(0.025)	0.055**
4		0.528***	(0.030)	0.226***	(0.025)	0.089***
Constant		2.833***	(0.109)	1.141***	(0.083)	0.224***

Table 7 (continued)

Dependent variable: Four-level self-rated health					
1 vs. 2–4		1–2 vs. 3–4		1–3 vs. 4	
Coeff	SE	Coeff	SE	Coeff	SE
N	24,759				
Rho	0.053***		(0.008)		

Source: Own calculations based on wave 1 to 4 of NIDS

The  $p$ -value from the Wald test of parallel lines assumption for the final model is 0.061, hence the null hypothesis that the parallel lines assumption cannot be rejected at the 1% and 5% significance levels

**Funding** Open access funding provided by University of KwaZulu-Natal.

## Declarations

**Conflict of Interests** The authors have no relevant financial or non-financial interests to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ardington, C., & Case, A. (2009). *Health: Analysis of the NIDS wave 1 dataset discussion paper no. 2*. Southern Africa Labour and Development Research Unit (SALDRU), Cape Town. Available from: <http://www.nids.uct.ac.za/publications/discussion-papers/wave-1-papers/93-nids-discussion-paper-no02/file>
- Ardington, C., & Gasealahwe, B. (2012). *Health: Analysis of the NIDS wave 1 and 2 datasets*. SALDRU Working Paper Number 80 / NIDS Discussion Paper 2012/3. Southern Africa Labour and Development Research Unit (SALDRU), Cape Town. Available from: <http://www.nids.uct.ac.za/documents/discussion-papers/wave-2/130-health>
- Ardington, C., & Gasealahwe, B. (2014). Mortality in South Africa: Socio-economic profile and association with self-reported health. *Development Southern Africa*, 31(1), 127–145.
- Bagod'Uva, T., van Doorslaer, E., Lindeboom, M., & O'Donnell, O. (2008). Does reporting heterogeneity bias the measurement of health disparities? *Health Economics*, 17(3), 351–375.
- Baidin, V., Gerry, C. J., & Kaneva, M. (2021). How self-rated is self-rated health? Exploring the role of individual and institutional factors in reporting heterogeneity in Russia. *Social Indicators Research*, 155(2), 675–696.
- Bailis, D. S., Segall, A., & Chipperfield, J. G. (2003). Two views of self-rated general health status. *Social Science & Medicine*, 56(2), 203–217.
- Baron, E. C., Davies, T., & Lund, C. (2017). Validation of the 10-item centre for epidemiological studies depression scale (CES-D-10) in Zulu, Xhosa and Afrikaans populations in South Africa. *BMC Psychiatry*, 17(1), 1–14.
- Beckett, S., Gould, W., Lillard, L., & Welch, F. (1988). The panel study of income dynamics after fourteen years: An evaluation. *Journal of Labor Economics*, 6(4), 472–492.
- Benyamini, Y., Blumstein, T., Lusky, A., & Modan, B. (2003). Gender differences in the self-rated health–mortality association: Is it poor self-rated health that predicts mortality or excellent self-rated health that predicts survival? *The Gerontologist*, 43(3), 396–405.
- Boyce, G., & Harris, G. (2011). A closer look at racial differences in the reporting of self-assessed health status and related concepts in South Africa. *Health SA Gesondheid (online)*, 16(1), 1–9.
- Brown, M., Daniels, R., de Villiers, L., Leibbrandt, M., & Woolard, I. (2012). *NIDS Wave 2 User Guide*. University of Cape Town.
- Brownson, R. C., Hoehner, C. M., Day, K., Forsyth, A., & Sallis, J. F. (2009). Measuring the built environment for physical activity: state of the science. *American Journal of Preventive Medicine*, 36(4), S99–S123.e112.
- Case, A., & Paxson, C. (2005). Sex differences in morbidity and mortality. *Demography*, 42(2), 189–214.

- Center for Behavioral Health Statistics and Quality. (2018). *2017 national survey on drug use and health: Methodological summary and definitions*. Rockville, MD: Substance Abuse and Mental Health Services Administration.
- Charasse-Poué, C., & Fournier, M. (2006). Health disparities between racial groups in South Africa: A decomposition analysis. *Social Science & Medicine*, *62*(11), 2897–2914.
- Chinhema, M., Brophy, T., Brown, M., Leibbrandt, M., Mlatsheni, C., Woolard, I. (2016). National income dynamics study panel user manual. *Cape Town: Southern Africa Labour and Development Research Unit*.
- Crossley, T. F., & Kennedy, S. (2002). The reliability of self-assessed health status. *Journal of Health Economics*, *21*(4), 643–658.
- Cutler, D.M., & Lleras-Muney, A. (2006). *Education and health: Evaluating theories and evidence*. NBER Working Paper No. 12352. Available from: <https://www.nber.org/papers/w12352>
- Cutler, D. M., Richardson, E., Keeler, T. E., & Staiger, D. (1997). Measuring the Health of the US Population. *Brookings Papers on Economic Activity, Microeconomics, 1997*, 217–282.
- De Villiers, L., Brown, M., Woolard, I., Daniels, R. & Leibbrandt, M. (eds.) (2013). *National Income Dynamics Study wave 3 user manual*. Cape Town: Southern Africa Labour and Development Research Unit (SALDRU). Available from: <http://www.nids.uct.ac.za/documents/wave-3-documents-and-questionnaires/157-nids-wave-3-user-manual/file>
- Denton, M., Prus, S., & Walters, V. (2004). Gender differences in health: A Canadian study of the psychosocial, structural and behavioural determinants of health. *Social Science & Medicine*, *58*(12), 2585–2600.
- Denton, M., & Walters, V. (1999). Gender differences in structural and behavioral determinants of health: An analysis of the social production of health. *Social Science & Medicine*, *48*(9), 1221–1235.
- DeSalvo, K. B., Blosner, N., Reynolds, K., He, J., & Muntner, P. (2006). Mortality Prediction with a Single General Self-Rated Health Question. *Journal of General Internal Medicine*, *21*(3), 267–275.
- Fallowfield, L. (2009). What is quality of life. *Health Economics*, *1*(8), 1–8.
- Ge, L., Ong, R., Yap, C. W., & Heng, B. H. (2019). Effects of chronic diseases on health-related quality of life and self-rated health among three adult age groups. *Nursing & Health Sciences*, *21*(2), 214–222.
- Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics*, *19*(3), 403–420.
- Guner, N., Kulikova, Y., & Llull, J. (2018). Reprint of: Marriage and health: Selection, protection, and assortative mating. *European Economic Review*, *109*, 162–190.
- Herman, A. A., Stein, D. J., Seedat, S., Heeringa, S. G., Moomal, H., & Williams, D. R. (2009). The South African Stress and Health (SASH) study: 12-month and lifetime prevalence of common mental disorders. *SAMJ: South African Medical Journal*, *99*(5), 339–344.
- Huisman, M., & Deeg, D. J. (2010). A commentary on Marja Jylhä's "What is self-rated health and why does it predict mortality? Towards a unified conceptual model"(69: 3, 2009, 307–316). *Social Science & Medicine*, *70*(5), 652–654.
- Idler, E. L. (2003). Discussion: Gender differences in self-rated health, in mortality, and in the relationship between the two. *The Gerontologist*, *43*(3), 372–375.
- Idler, E., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, *38*(1), 21–37.
- Jelsma, J., & Ferguson, G. (2004). The determinants of self-reported health-related quality of life in a culturally and socially diverse South African community. *Bulletin of the World Health Organization*, *82*(3), 206–212.
- Jürges, H. (2007). True health vs response styles: Exploring cross-country differences in self-reported health. *Health Economics*, *16*(2), 163–178.
- Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science & Medicine*, *69*(3), 307–316.
- Kaplan, G., & Baron-Epel, O. (2003). What lies behind the subjective evaluation of health status? *Social Science & Medicine*, *56*(8), 1669–1676.
- Kaplan, R. M. (2003). The significance of quality of life in health care. *Quality of Life Research*, *12*, 3–16.
- Kawachi, I., & Berkman, L. F. (2003). *Neighborhoods and health*. Oxford University Press.
- Kerkhofs, M., & Lindeboom, M. (1995). Subjective health measures and state dependent reporting errors. *Health Economics*, *4*(3), 221–235.

- Khan, H. T., & Raeside, R. (2014). Between country variations in self-rated-health and associations with the quality of life of older people: Evidence from the global ageing survey. *Applied Research in Quality of Life*, 9, 923–949.
- Kiecolt-Glaser, J. K., & Newton, T. L. (2001). Marriage and health: His and hers. *Psychological Bulletin*, 127(4), 472.
- Lehohla, P. (2013). *Use of health facilities and levels of selected health conditions in South Africa: Findings from the General Household Survey, 2011*. Pretoria: Statistics South Africa.
- Lindeboom, M., & van Doorslaer, E. (2004). Cut-point shift and index shift in self-reported health. *Journal of Health Economics*, 23(6), 1083–1099.
- Macintyre, S., Hunt, K., & Sweeting, H. (1996). Gender differences in health: Are things really as simple as they seem? *Social Science & Medicine*, 42(4), 617–624.
- McLeod, H. (2012). The role of risk adjustment in the equitable financing of National Health Insurance in South Africa. *Development Southern Africa*, 29(5), 636–656.
- Paul, P., & Valtonen, H. (2016). Inequalities in perceived health in the Russian Federation, 1994–2012. *BMC Public Health*, 16(1), 1–12.
- Pfarr, C., Schmid, A., & Schneider, U. (2011). Estimating ordered categorical variables using panel data: A generalized ordered probit model with an autofit procedure. *Journal of Economics and Econometrics*, 54(1), 7–23.
- Pfarr, C., Schmid, A., & Schneider, U. (2012). Reporting heterogeneity in self-assessed health among elderly Europeans. *Health Economics Review*, 2(1), 1–14.
- Posel, D. (2010). Households and labour migration in post-apartheid South Africa. *Studies in Economics and Econometrics*, 34(3), 129–141.
- Posel, D. & Hall, K. (2021) "Chapter 37: The economics of households in South Africa". In A. Oqubay, F. Tregenna, & I. Valodia (Eds.) *The Oxford handbook of the South African economy*. Oxford Academic
- Radloff, L. S. (1977). The CES-D scale a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3), 385–401.
- Rossouw, L., Bagod'Uva, T., & Van Doorslaer, E. (2018). Poor health reporting? Using anchoring vignettes to uncover health disparities by wealth and race. *Demography*, 55(5), 1935–1956.
- Schneider, U., Pfarr, C., Schneider, B. S., & Ulrich, V. (2012). I feel good! Gender differences and reporting heterogeneity in self-assessed health. *The European Journal of Health Economics*, 13(3), 251–265.
- Sen, A. (2002). Health: Perception versus observation: Self reported morbidity has severe limitations and can be extremely misleading. *British Medical Journal*, 324(7342), 860.
- Smith, K. W., Avis, N. E., & Assmann, S. F. (1999). Distinguishing between quality of life and health status in quality of life research: A meta-analysis. *Quality of Life Research*, 8, 447–459.
- Southern Africa Labour and Development Research Unit (SALDRU) (2008). *National Income Dynamics Study 2008, wave 1* [dataset]. Version 7.0.0. Pretoria: SA Presidency [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2018. Cape Town: DataFirst [distributor], 2018. <https://doi.org/10.25828/e7w9-m033>
- Southern Africa Labour and Development Research Unit (SALDRU) (2010). *National Income Dynamics Study 2010-2011, wave 2* [dataset]. Version 4.0.0. Pretoria: SA Presidency [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2018. Cape Town: DataFirst [distributor], 2018. <https://doi.org/10.25828/j1h1-5m16>
- Southern Africa Labour and Development Research Unit (SALDRU) (2012). *National Income Dynamics Study 2012, wave 3* [dataset]. Version 3.0.0. Pretoria: SA Presidency [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2018. Cape Town: DataFirst [distributor], 2018. <https://doi.org/10.25828/7pgq-q106>
- Southern Africa Labour and Development Research Unit (SALDRU) (2014). *National Income Dynamics Study 2014-2015, wave 4* [dataset]. Version 2.0.0. Pretoria: Department of Planning, Monitoring, and Evaluation [funding agency]. Cape Town: Southern Africa Labour and Development Research Unit [implementer], 2018. Cape Town: DataFirst [distributor], 2018. <https://doi.org/10.25828/f4ws-8a78>
- Tubeuf, S. (2009). Income-related health inequalities in France in 2004: Decomposition and explanations. *Revue D'épidémiologie Et De Santé Publique*, 57(5), e34–e43.
- van Doorslaer, E., & Jones, A. M. (2003). Inequalities in self-reported health: Validation of a new approach to measurement. *Journal of Health Economics*, 22(1), 61–87.

- Wittenberg, M. (2013). The Weight of Success: The Body Mass Index and Economic Well-Being in Southern Africa. *Review of Income and Wealth*, 59(S1), S62–S83.
- World Health Organization (WHO). (1948). *Constitution of the World Health Organization*. World Health Organization. Available from: <https://apps.who.int/gb/bd/PDF/bd47/EN/constitution-en.pdf>
- World Health Organization (WHO). (2000). Obesity: preventing and managing the global epidemic. Report of a WHO consultation. *World Health Organization Technical Report Series*, 894, i–xii, 1–253. Available from: [http://whqlibdoc.who.int/trs/WHO\\_TRS\\_894.pdf](http://whqlibdoc.who.int/trs/WHO_TRS_894.pdf)
- World Health Organization (WHO). (2010). *Global recommendations on physical activity for health*. World Health Organization. Available from: <https://apps.who.int/iris/handle/10665/44399>
- Wu, S., Wang, R., Zhao, Y., Ma, X., Wu, M., Yan, X., & He, J. (2013). The relationship between self-rated health and objective health status: A population-based study. *BMC Public Health*, 13(1), 1–9.
- Ziebarth, N. (2010). Measurement of health, health inequality, and reporting heterogeneity. *Social Science & Medicine*, 71(1), 116–124.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.