Adjusting for Misclassification in the Analysis of Self-Rated Health: Evidence from Italy

Eleonora Trappolini, Maria Felice Arezzo, Giuseppina Guagnano
Sapienza University of Rome

Introduction and background

Evaluating population's health conditions has always been a crucial topic in demographic research. Recently, the interest towards population health was renewed as a consequence of the ageing process, which characterises most high-income countries.

In survey research on health there are a number of health indices to measure individuals' health status (Lundberg & Manderbacka, 1996). An alternative approach involves inviting survey participants to answer to a seemingly straightforward question 'How is your health in general?', in which respondents are asked to self-assess their health as 'very good,' 'good,' 'fair,' 'bad,' 'very bad.' This type of health measure, known as self-rated health (SRH), is a multifaceted tool commonly used for the evaluation of individuals' health and it is also widely adopted due to its recognised substantial predictive and concurrent validity in the field of epidemiology and population health research. This validity is demonstrated by strong correlations with future health outcomes, including mortality and a range of morbidity, disability, and healthcare service utilisation indicators (Idler & Benyamini, 1997; Jylhä, 2009). Despite its extensive use in assessing health and its ability to predict adverse health outcomes, SRH as an indicator of individuals' true health status is met with skepticism. This skepticism stems from the broader debate on the contrast between subjective perceptions and objective observations. Research suggests that self-reports of health and illness can be influenced by social experiences, potentially leading to inaccuracies, particularly among socially disadvantaged individuals from low- and middle-income backgrounds (Sen, 2002).

Measurement reliability is crucial for ensuring validity. Insufficient reliability can introduce measurement errors, potentially biassing analytical results, especially when the construct serves as an independent or dependent variable. In explaining an unreliable measure of SRH through an ordered regression model, it is known from the statistical literature that the estimators of the regression coefficients are inconsistent (Hausman, 1998). Therefore, to have consistent estimates of the regression parameters it is of paramount importance to rely upon an appropriate statistical methodology.

Moreover, the reliability levels may vary across different population subgroups, potentially affecting empirical comparisons. This is particularly significant when assessing health disparities, given the frequent use of SRH as a measure (Zajacova & Dowd, 2011).

SRH, as a subjective health measure, distinguishes itself from most other health indicators due to the absence of formal, agreed-upon rules or definitions guiding the process of how individuals rate their health. As Jylhä (2009) suggests, SRH evaluation inherently requires the processing of information, the interpretation of meanings, and a process of selection. Despite being an individual, subjective process, it operates within the confines of a specific social and cultural context and draws upon the conceptual resources and patterns of representation provided by that environment.

In Juerges' work (2007), an attempt was made to distinguish the impact of 'true health' from the impact of response styles in cross-country variations. This was achieved by using other health metrics present in the study to 'adjust' the self-ratings. However, this intriguing approach has its limitations as it fails to account for the unmeasured aspects of health not covered by 'objective' indicators and the intricacies of the evaluation processes (Jylhä, 2009).

There are also other studies that have explored the reliability of SRH (Crossley & Kennedy, 2002; Lundberg & Manderbacka, 1996). These studies revealed notable variations in SRH ratings when the health status questionnaire was administered on two separate occasions. However, it is important to acknowledge that perceptions and reporting of subjective health can differ significantly across countries (Salomon et al., 2004). In another study by Zajacova and Dowd (2011), the authors assessed the test-retest reliability of general SRH within a nationally representative sample of U.S. adults. They collected responses from two interviews conducted approximately one month apart and found that nearly 40% of respondents changed their health rating, indicating moderate test-retest reliability of SRH. They also examined reliability across key sociodemographic variables, including age, gender, race, and educational attainment.

Furthermore, other studies demonstrate the reliability of this health measure, suggesting that in the health status ratings, it is common to find respondents clustered in middle-range options rather than in extreme ones (Bowling & Windor, 2008). For instance, participants prefer 'good' over 'very good' to potentially conceal some reservations about their health status (Fakhoury et al., 2021), or they may prefer 'good' over 'fair' as the latter is perceived as too negative (Perneger et al., 2007).

Drawing upon this literature, the aim of this study is to propose an approach for adjusting the estimates of the regression coefficients and, consequently, the estimates of the probabilities associated with each SRH level. The approach will be employed to analyse the relationship between SRH and a set of typical predictors (e.g., educational level, marital status) among the elderly in Italy.

Data and methods

For our study, we used the 2019 Italian data from the European Health Interview Survey (IT-EHIS, wave 3) carried out by the National Institute of Statistics (Istat). The IT-EHIS survey collects information on the health status, healthcare provision, health determinants and socioeconomic condition of the Italian population. The sample includes data on nearly 46,000 individuals aged 15 and above, residing in private dwellings. Given our study's focus on the health of the elderly population, we restricted our sample to individuals aged 50 and above, excluding those with missing information on health related variables, thus, the final sample size is of 25,672 individuals.

The dependent variable is SRH, derived from the question 'How is your health in general?' with five possible answers: 'very good', 'good', 'fair', 'bad', 'very bad'. We categorised this variable into three groups: 0 for 'very good' and 'good'; 1 for 'fair'; and 2 for 'bad' and 'very bad'.

Independent variables include age groups ('50-54', '55-59', '60-64', '65-69', '70-74', '75+'), marital status ('single', 'married', and 'divorced or widow'), education ('no education or primary', 'lower secondary', 'upper secondary', 'tertiary'), employment status ('retired', 'employed', 'unemployed', 'housekeeper' 'inactive'), migration background ('native', 'foreign-born'), and macro-area of residence in Italy ('North', 'Centre', 'South') (Table 1).

Table 1 - Sample characteristics and distribution of the outcome by sex

		Men	**Women**
Age groups			
	50-54	17.82	16.20
	55-59	16.49	15.03
	60-64	14.89	13.88
	65-69	13.91	13.19
	70-74	12.99	12.26
	75+	23.90	29.44
Marital status			

Single	11.39	8.51
Married	76.02	58.77
Divorced/Widow	12.59	32.72
Education		
No education/Primary	22.82	33.16
Lower secondary	34.27	28.94
Upper secondary	30.80	27.72
Tertiary	12.12	10.17
Employment status		
Retired	52.39	35.24
Employed	39.58	24.27
Unemployed	4.67	3.07
Housekeeper	0.18	33.66
Inactive	3.19	3.76
Migration background		
Native	95.87	93.63
Foreign-born	4.13	6.37
Macro-area of residence		
North	44.78	44.55
Centre	19.44	19.36
South	35.77	36.09
Outcome		
Self-rated health		
Very good/Good	58.18	49.52
Fair	30.76	34.71
Poor/Very poor	11.06	15.77
N. observations	*11,810*	*13,862*

Note: The Table shows percentages and should be read in columns.
Source: Authors' elaboration on IT-EHIS (wave 3) data.

Methods

In many disciplines data are frequently misclassified. In the statistical literature, there are several models that address misclassification in binary regression (Hausman, 1998; Naranjo et al., 2014; Arezzo and Guagnano, 2019; Arezzo et al., 2023), but there are few models that consider measurement error for polychotomous responses. Misclassification of a dependent variable Y means that an observation with a true value *j* is observed as *k*. This mistake could easily happen, for example, during an interview if the respondent misunderstands the question or the interviewer simply checks the wrong box.

Ignoring the presence of misclassification is not trivial; in fact, when traditional estimation methods (like ordered logit or probit) are used in discrete-choice contexts with a misclassified dependent variable, the resulting estimates are inconsistent.

To obtain consistent estimates of the regression coefficients β, we need to explicitly take into account the misclassification probabilities $\Pr(Y^O = j|Y^T = k) = \alpha_{jk}$. In particular, the probability of observing *j* conditionally on a set of independent variables is:

$$\Pr(Y^O = j|X) = \sum_{k=1}^{K} a_{jk} \Pr(Y^T = k|X) \qquad (1)$$

where K are the categories of the response variable. We then model $\Pr(Y^T = k|X)$ via a cumulative ordered logit model as:

$$\Pr(Y^T \leq k|X) = \Lambda(\gamma_k - X\beta) \qquad (2)$$

where $\Lambda(.)$ is the logistic cumulative density function.
The vector of parameters $\theta = (\beta, \gamma, \alpha)$ is estimated through likelihood maximisation.

For the purpose of our study, we conducted two separate analyses. We first fit a standard ordered logit model and then an adjusted one, comparing the results.
All analyses are stratified by sex.

Expected results

In line with literature on misclassification (Neuhaus, 1999; Liu and Zhang, 2017), we expect the association between the outcome variable and the independent variables will be weakened in the naive model in case of misclassification.
Additionally, consistent with previous findings about SRH, we expect higher levels of education levels to correlate with better SRH as well as being in a stable relationship to have a protective effect.
This study addresses the subjectivity and potential misclassification in SRH. By applying adjusted models, we aim to obtain more accurate estimates in understanding SRH and its relationship with various predictors, contributing to better population health assessment.

References

Arezzo, M.F. & Guagnano, G. (2019). Misclassification in binary choice models with sample selection. *Econometrics,* 7(3), 32.

Arezzo, MF, Arima, S. & Guagnano, G. (2023). A two-part measurement error model to estimate participation in undeclared work and related earnings. *Statistical Modelling.*

Bowling, A., & Windsor, J. (2008). The effects of question order and response-choice on self-rated health status in the English Longitudinal Study of Ageing (ELSA). *Journal of the Epidemiology and Community Health*, 62, 81–85.

Crossley, T. F., & Kennedy, S. (2002). The reliability of self-assessed health status. *Journal of Health Economics*, 21(4), 643-658.

Fakhoury, J., Burton-Jeangros, C., Guessous, I., Consoli, L., Duvoisin, A., & Jackson, Y. (2021). Self-rated health among undocumented and newly regularized migrants in Geneva: a cross-sectional study. *BMC Public Health*, 21, 1198.

Hausman, J., Abrevaya, J. & Scott-Morton, F. (1998). Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics*, 87(2):239 – 269.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior*, 21-37.

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.

Jürges, H. (2007). True health vs response styles: exploring cross-country differences in self-reported health. *Health Economics*, 16(2), 163-178.

Liu, H. & Zhang, Z. (2017). Logistic regression with misclassification in binary outcome variables: a method and software. *Behaviormetrika*, 44:447–476

Lundberg, O., & Manderbacka, K. (1996). Assessing reliability of a measure of self-rated health. *Scandinavian Journal of Social Medicine*, 24(3), 218-224.

Naranjo, L., Martın, J., Perez, C.J. & Rufo, M.J. (2014). Addressing misclassification for binary data: probit and t-link regressions, Journal of Statistical Computation and Simulation 84, 2187–2213.

Neuhaus, J. M. (1999). Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*, 86 (4), 843–855.

Perneger, T. V., Gayet-Ageron, A, Courvoisier, D.S., Agoritsas, T., & Cullati, S. (2013). Self-rated health: analysis of distances and transitions between response options. *Quality of Life Research*, 22, 2761–2768.

Salomon, J. A., Tandon, A., & Murray, C. J. (2004). Comparability of self-rated health: cross sectional multi-country survey using anchoring vignettes. *BMJ*, 328(7434), 258.

Sen, A. (2002). Health: perception versus observation: self-reported morbidity has severe limitations and can be extremely misleading. *BMJ*, 324(7342), 860-861.

Zajacova, A., & Dowd, J. B. (2011). Reliability of self-rated health in US adults. *American Journal of Epidemiology*, 174(8), 977-983.