**Agne Suziedelyte, School of Economics, University of New South Wales**[1]
**Meliyanni Johar, Economics Discipline Group, University of Technology Sydney**

## Abstract

We investigate the common assumption in applied research that reporting errors are negligible in variables where there is no clear incentive for misreporting. Using major medical operations, we find high misreporting rates, but the coefficients of their predictors remain unbiased.

**Keywords**: measurement error, health, socio economic status

**JEL Codes**: C1, I1, I2

## Acknowledgement

This research is funded by an ARC Discovery Project grant (DP110100729). It uses data from the 45 and Up Study, which is managed by the Sax Institute in collaboration with major partner Cancer Council New South Wales, and partners the Heart Foundation (NSW Division); NSW Ministry of Health; *beyondblue*; Ageing, Disability and Home Care, NSW Family and Community Services; Australian Red Cross Blood Service; and UnitingCare Ageing. This project was undertaken by the University of Technology Sydney and utilised PBS data supplied by the Department of Human Services and linked to the 45 and Up Study by the Sax Institute. Data linkage for the hospital admission data used in the project was undertaken by the Centre for Health Record Linkage. The project has ethics approval from the NSW Population and Health Services Research Ethics Committee and the Department of Health and Ageing Departmental Ethics Committee. The study's findings are those of the authors' and do not necessarily represent the views of the Department of Health and Ageing, or the Department of Human Services.

## Introduction

The ever-growing acknowledgement of non-random measurement error in self-reported data has motivated researchers to investigate its impact (e.g., Johnston et al., 2009; Akee, 2011; Erickson and Whited, 2000; Forbes, 2000; Dwyer and Mitchell, 1999) and to improve techniques to deal with it in empirical models (Bound et al., 2001; Black et al., 2000; Brownstone and Valletta, 1996). The measurement error literature has focused on variables where misreporting may be perceived to provide an advantage, real or psychological, for the respondent. For example, studies have examined measurement errors in self-reported earnings

[1] Corresponding author. Australian School of Business Building, Level 4, 447, UNSW Sydney NSW 2052, Australia, phone +61 2 9385 2660, e-mail: agne@unsw.edu.au

(Akee 2011; Brownstone and Valletta, 1996); disability status (Kreider and Pepper, 2007); grades, class ranks and test scores (Kuncel et al., 2005); height, weight and body mass index (Gorber 2007); sexual behaviour (Tennekoon and Rosenman, 2013); and subjective health or wellbeing (Bago d'Uva et al., 2006; Dwyer and Mitchell, 1999). Labour market outcomes, wealth and income are often misreported to gain a tax or subsidy advantage; principals and teachers may inflate their school's achievement scores in order to improve reputation; due to stigma associated with undesirable body shape, survey respondents often misreport their weight and height; subjective variables such as self-rated general health or well-being may not reflect the true health and well-being conditions, because the answers to these questions are affected by adaptation bias and cultural norm. For other variables, there is no clear incentive for the survey respondents to be untruthful and/or they are deemed to be objective. These variables may also be reported with errors, but they are often assumed to be harmless and random (conditional on basic demographics). The purpose of this study is to confirm such assumption, using the case of life-changing medical operations.

In contrast to previous studies, we investigate reporting errors in variables that might be expected to have small, most likely random errors. If they are found to be truly random and therefore harmless to subsequent analysis then we can have confidence in relying on the self-reports of such variables. We use the case of major medical operations that are likely to be memorable. We cross-check the self-reports of operations with contemporaneous hospital administrative records. To examine the consequence of ignoring a reporting error, we compare the socio economic status (SES)-health gradient estimated based on the self-reported and administrative data.

Our study is related to Johnston et al (2009) who show that reliance on self-reported chronic conditions can lead to underestimation of income-related inequalities in health because low income individuals are more likely than high income individuals to under-report bad health. As a result, the size of the income-health gradient using measured health is much stronger than that implied by self-reported health. Their study uses hypertension as a measure of health and finds that over 85% of individuals measured by a trained nurse to have hypertension do not indicate in the survey to have hypertension. One issue with the study of Johnston et al (2009) is that disparity in hypertension rates may reflect infrequent health checks, rather than reporting error. Moreover, the respondents may have been prescribed hypertension medication and believe that their blood pressure has returned to normal. Our use of major, life-changing operations does not suffer from these drawbacks.

## Data and Method

Our analysis sample is based on 241,138 non-institutionalised individuals who participated in the 45 and Up Study[2] in the state of New South Wales (NSW), Australia, fielded in stages during 2006-2009; NSW has a population of about 7.3 million with 39% aged over 45. In this

[2] For details see http://www.saxinstitute.org.au/ . Participants were randomly selected from the central database, which contains everyone who has ever used health services in Australia, for this population age group. The survey was done in stages, but the bulk of it (about 80%) was collected in 2008. This variation in survey year is due to sampling process rather than the choice of respondents.

survey, respondents were prompted with a list of operations and asked whether they had any of these operations, and for each of them, the age at the most recent operation. The 45 and Up Study is linked to the Admitted Patient Data Collection (APDC) consisting of all admissions in NSW hospitals during 2000-2009.[3] This data linkage allows us to cross-check the consistency of the self-reported history of operations with the hospital records during this period. From the operations listed in the survey, we select four operations that are performed in a hospital setting and specifically defined so that matching them in the administrative data is straightforward. For example, it is difficult to match "heart operation", because we cannot tell from the survey question which procedures should be included in the definition of "heart operation" in the administrative data. The four operations are knee replacement, removal of gall bladder, removal of prostate, and hysterectomy.

So that all respondents have an equal cross-check window, we focus on the hospital records in the last 5 years from the survey date. To be consistent, using the survey data, we create binary variables that indicate whether or not a respondent reported an operation in the last 5 years. Imposing the 5 years window, we define four possible cases: (1) an operation found both in the survey and hospital data (true positive); (2) an operation not found in the survey and hospital data (true negative); (3) an operation found in the administrative data but not in the survey data (false negative or under-reporting); and (4) an operation found in the survey data but not in the administrative data (false positive or over-reporting).

Our analysis has three stages. In the first stage for each of the above mentioned operations, we present the under-reporting rates, calculated as the proportion of operations in the hospital records that are not found in the survey data, and the over-reporting rates, calculated as the proportion of operations in the survey data that are not found in the hospital records. The under-reporting rates are due to respondents either: (i) not reporting the operation at all or (ii) misreporting the timing of the operation. Since we are only looking at operations in the last 5 years, it is unlikely that respondents would forget having them, especially since the selected operations are life-changing, occurring only once in a lifetime and/or involving major interventions. But there could be other reasons why respondents did not report an operation, including the question being considered too sensitive and lack of attention or medical knowledge (comprehension) in filling in the survey. On the other hand, the over-reporting rates may be also explained by the limited coverage of the administrative data. For instance, operations overseas will not be captured in the data, although we can expect them to be rare given that Australian public hospitals provide these operations for free. Operations performed in other Australian states are also outside the data range, but studies suggests that incentives to travel inter-state are small (e.g., Johar et al., 2012).

In the second stage, we investigate whether the propensity of misreporting (over or under) can be explained by individual characteristics, capturing variation in memory capabilities, preferences for health, and medical knowledge. R-squared statistics from OLS regressions are used as measures of explanatory power. We reveal the proportion of the variation in reporting

[3] The data linkage is undertaken by the Centre for Health Records Linkage and the linked, de-identified data is released under ethics approval.

error that is explained by individual characteristics such as age, sex, ethnicity, marital status, education, income, employment, housing, neighborhood characteristics, and family health history. We also investigate whether own health variables have additional explanatory power over and above the other individual background variables.

The last stage assesses the implications of reliance on survey data. We run probit regressions of the propensity of having each of these operations using first survey then hospital data. We pay specific attention to the slope coefficients of SES measures: education, income and employment. If the measurement error is random, we would find a consistent SES-health gradient, regardless of the source of data.

In addition, we also look at hypertension, which unlike operations, may be harder for an individual to notice, because it is often asymptomatic. Previous literature finds that misreporting in such hard-to-observe variables correlates with SES (Johnston et al., 2009). We verify self-reported diagnosis of hypertension using administrative data on claims for anti-hypertensive prescription drugs and hypertension diagnoses in hospital admission records in the next 12 months from the survey date.

**Results**

The first two columns of Table 1 present the incidence of the selected operations and hypertension by each data source. The incidence of most of the operations is higher in the survey data than in the hospital records, suggesting that respondents are more likely to over-report than under-report the operations. On the other hand, hypertension tends to be slightly over-reported. Columns 3 and 4 show the extent of the under- and over-reporting. As an example, there are 1,294 self-reports of knee replacements that are not found in the hospital data, giving an over-reporting rate of 26.4% (1294/4896), and there are 686 hospital admissions for knee replacement that are not reported in the survey, giving an under-reporting rate of 16.0% (686/4288). The over-reporting rates are surprisingly high (24.3%-29.1%) and are similar for all the operations, whereas the under-reporting rates are on average lower and vary from 8.8% for hysterectomies to 32.3% for prostate removal operations. These high over-reporting rates may be partly explained by respondents mistakenly reporting operations that should not be included in the definition of a particular surgery. For example, when other procedures on prostate are added to the definition of prostate removal surgery in administrative data, the false positive rate goes down from 24.3% to 20.8%. The over- and under-reporting rates of hypertension are slightly higher than the average misreporting rates of the operations[4].

Next, we investigate whether the under-reporting rates are driven by misreporting of age at the operation. For this purpose, we separate the operations found in the hospital records but not in the survey into those not reported in the survey at all (column 5) and those reported but with an error in age, that is, outside the 5 year period (column 6). The results show that misreporting

[4] Note that the pharmaceutical administrative data record only claims that are eligible for a subsidy (i.e., drugs that cost above a co-payment or if a claimant has reached the Medicare's safety net threshold). If we restrict the sample to a small sub-sample of health care concession card holders, who are more heavily subsidized and for whom the claims are accurately recorded, the over-reporting rate of hypertension is lower (7.24%) while the underreporting rate is comparable to the full sample at 29.7%.

of age at the operation is not driving the under-reporting rates, except for knee replacements. Most of the under-reporting rates are due to not reporting the surgery in the survey at all.

To further check how accurately individuals recall the timing of their operations, we use the respondents who had an operation in the 5 year period before the survey date and reported this operation in the survey, possibly with an error in age, which can vary from 0 to 4 years. Table 2 presents the distribution of the discrepancy in age by operation. These results also show that age misreporting is not common, as most of the respondents either report the age accurately or misreport age by only one year.

Next, we investigate whether or not reporting error in the operations and hypertension status can be explained by individual characteristics. For each of these variables, we create a binary variable that takes the value one if a respondent over- or under-reported an operation or hypertension status and the value zero otherwise. Then, we regress these variables on individual background characteristics and health measures. Table 3 reports the R-squared statistics for these OLS regressions. The results show that an extensive list of individual background characteristics (66 variables) can explain only a small proportion of the variation in the reporting errors. Adding own health measures improves R-squared statistics, but most of the measurement error remains unexplained. We find that the SES variables predict the misreporting in hypertension better than the misreporting in the operations (F-statistic for the joint significance test is 51.5 for hypertension and 1.0 – 6.3 for the operations).

Finally, we analyse whether the observed misreporting behaviour affects the estimation of the impact of SES on health, as measured by the above mentioned operations and hypertension status[5]. Table 4 presents the estimated average partial effects of education, income, and employment on the probabilities of having these health problems, first using the survey data and then the hospital records. The regressions control for the individual background characteristics. In the case of the operations, the estimated coefficients are very similar irrespective of the data source used, consistent with the earlier finding that the reporting error is largely unexplained by SES. In contrast, the effects of SES on hypertension differ substantially depending on the data source used. Specifically, these effects are under-estimated when the survey data is used to measure hypertension, which is consistent with Johnston et al's (2009) results. As in the previous studies, we find positive effects of SES on health (e.g., Lleras-Muney, 2005; Currie, 2009; Johnston et al., 2009; Conti et al., 2010).

## Concluding remarks

Without restricting our study to a health audience, we carefully consider the measurement error in major health operations. We choose these variables because they are objective, memorable events, which intuitively should be accurately reported. In addition, unlike previous studies, which use externally measured variables (by interviewer, nurse, etc) as the true information to

[5] Under the well-functioning Australian universal health care system, all sick individuals have access to medical treatment so it is reasonable to assume that heath care utilisations capture health state. Nonetheless, it is possible that operations also capture non-health factors.

compare against the self-reported data, we cross-check the self-reports with the linked contemporaneous administrative data. This avoids yet another potential measurement error, for example, due to the gaps between the survey and measurement dates.

Our results provide comfort to many applied researchers who rely on survey data for analysis and suggest that we can rely on survey variables that are objective in nature. While the measurement errors are large, they do not create bias in the slope estimates when the survey variables are used as dependent variables. Previously, researchers had assumed (and were hopeful) that correction for potential reporting errors is not necessary for these types of objective variables, but until now this assumption has not yet been put to a real empirical test. On the other hand, our findings suggest that measurement error in harder-to-observe variables may lead to substantially biased slope estimates.

## References

Akee R. 2011. Errors in self-reported earnings: The role of previous earnings volatility and individual characteristics. Journal of Development Economics 96(2): 409–421.

Black DA, Berger MC, Scott FA. 2000. Bounding parameter estimates with nonclassical measurement error. Journal of the American Statistical Association 95(451): 739-748.

Bound J, Brown C, Mathiowetz N. Measurement error in survey data. In J.J. Heckman & E.E. Leamer (ed.), 2001. "Handbook of Econometrics," Handbook of Econometrics, Elsevier, edition 1, volume 5, number 5, 5-59.

Brownstone D, Valletta RG. 1996. Modeling earnings measurement error: A multiple imputation approach. The Review of Economics and Statistics 78(4): 705-717.

Butler JS, Burkhauser RV, Mitchell JM, Pincus TP. 1987. Measurement error in self-reported health variables. The Review of Economics and Statistics 69(4): 644-650.

Conti, Gabriella, James J. Heckman, and Sergio Urzua. 2010. \The Education-Health Gradient." American Economic Review Papers and Proceedings 100: 234-238.

Currie J. 2009. Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. Journal of Economic Literature 47(1): 87-122.

Dwyer DS, Mitchell OS. 1999. Health problems as determinants of retirement: Are self-rated measures endogenous? Journal of Health Economics 18(2): 173-193.

Erickson T, Whited TM. 2000. Measurement Error and the Relationship between Investment and *q*. Journal of Political Economy 108(5): 1027-1057.

Forbes KJ. 2000. A Reassessment of the Relationship between Inequality and Growth. American Economic Review 90(4): 869-887.

Gorber SC, Tremblay M, Moher D, Gorber B. 2007. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. Obesity Reviews 8: 307–326.

Johar M, Jones G, Keane M, Savage E, Stavrunova O. 2012. Geographic Differences in Hospital Waiting Times. Economic Record 88(281): 165-181.

David W. Johnston, Carol Propper, Michael A. Shields, Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient, Journal of Health Economics, 28(3): 540-552.

Kreider B, John VP. 2007. Disability and employment: reevaluating the evidence in light of reporting errors. Journal of the American Statistical Association 102(478): 432-441.

Lleras-Muney A. 2005. The relationship between education and adult mortality in the United States. The Review of Economic Studies 72(1): 189-221.

Kuncel N, Credé M, Thomas LL. 2005. The Validity of Self-Reported Grade Point Averages, Class Ranks, and Test Scores: A Meta-Analysis and Review of the Literature. Review of Educational Research 75(1): 63-82.

Tennekoon V, Rosenman R. 2013. ‘Behold, a virgin is with HIV!’ Misreporting sexual behavior among infected adolescents. Health Economics. doi: 10.1002/hec.2918

Table 1. Incidence, under-reporting and over-reporting of operations

| | Incidence | | Over-reporting | Under-reporting | | |
|---|---|---|---|---|---|---|
| | Survey | Admin. | Total | Total | Misreport status | Misreport age |
| Prostate | 3,608 | 4,034 | 24.3% | 32.3% | 23.2% | 9.0% |
| Hysterectomy | 3,931 | 3,057 | 29.1% | 8.8% | 5.1% | 3.7% |
| Knee | 4,896 | 4,288 | 26.4% | 16.0% | 4.9% | 11.1% |
| Gall bladder | 4,770 | 4,207 | 27.1% | 17.4% | 12.3% | 5.0% |
| Hypertension | 84,611 | 80,341 | 31.6% | 27.9% | - | - |

Table 2. Distribution of error in age at the operation

| | No error | 1 year | 2 years | 3 years | 4 years | Sample size |
|---|---|---|---|---|---|---|
| Prostate | 63.54% | 30.82% | 4.69% | 0.92% | 0.04% | 2,732 |
| Hysterectomy | 64.35% | 31.67% | 3.52% | 0.43% | 0.04% | 2,788 |
| Knee | 61.83% | 32.18% | 4.66% | 0.94% | 0.39% | 3,602 |
| Gall bladder | 57.09% | 34.86% | 6.70% | 1.24% | 0.09% | 3,477 |

Table 3. Proportion of reporting error explained by individual characteristics

| | Background | Health | Mean error |
|---|---|---|---|
| Prostate | 0.016 | 0.034 | 0.019 |
| Hysterectomy | 0.002 | 0.005 | 0.011 |
| Knee | 0.006 | 0.012 | 0.008 |
| Gall bladder | 0.002 | 0.003 | 0.008 |
| Hypertension | 0.018 | 0.043 | 0.205 |

Notes: For operations, sample sizes are 111,983 for males, 129,155 for females, and 241,138 overall. For hypertension, sample size is 239,467. The reported figures in columns 1 and 2 are R-squared statistics from OLS regressions. In column 1, the model includes background characteristics: demographic and socio-economic variables, neighbourhood characteristics and family health history. The model in column 2 additionally includes health measures. Column 3 reports mean measurement error in the sample.

Table 4. Variation in the estimated effects of socio-economic status variables by data source

| | Prostate | | Hysterectomy | | Knee | | Gall bladder | | Hypertension | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Survey | Admin. | Survey | Admin. | Survey | Admin. | Survey | Admin. | Survey | Admin. |
| *Education [a]:* | | | | | | | | | | |
| High school (year 12) | -0.007*** | -0.007*** | -0.003* | -0.003* | -0.001 | -0.002** | -0.002** | -0.002*** | -0.001 | -0.012*** |
| | (0.002) | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.003) | (0.003) |
| University | 0.000 | -0.002 | -0.003*** | -0.003*** | -0.006*** | -0.005*** | -0.003*** | -0.004*** | -0.034*** | -0.043*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.002) | (0.002) |
| *Household income [b]:* | | | | | | | | | | |
| $30,000-$49,999 pa | -0.002 | -0.002 | 0.001 | 0.000 | -0.002*** | -0.001 | -0.004*** | -0.004*** | -0.015*** | -0.058*** |
| | (0.002) | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.003) | (0.002) |
| $50,000-$69,999 pa | -0.001 | -0.002 | 0.000 | -0.003* | -0.003*** | -0.002* | -0.003*** | -0.002** | -0.016*** | -0.098*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) | (0.004) | (0.003) |
| >$69,999 pa | -0.001 | -0.001 | -0.001 | -0.002 | -0.003*** | -0.002* | -0.005*** | -0.005*** | -0.014*** | -0.126*** |
| | (0.002) | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.003) | (0.003) |
| Employed | -0.001 | -0.001 | -0.000 | 0.000 | -0.007*** | -0.006*** | -0.003*** | -0.003*** | -0.040*** | -0.107*** |
| | (0.001) | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.003) | (0.002) |
| Pseudo R2 | 0.072 | 0.085 | 0.028 | 0.027 | 0.086 | 0.095 | 0.018 | 0.016 | 0.115 | 0.261 |
| Sample size | 111,983 | 111,983 | 129,155 | 129,155 | 241,138 | 241,138 | 241,138 | 241,138 | 239,467 | 239,467 |
| Mean | 0.032 | 0.036 | 0.030 | 0.024 | 0.020 | 0.018 | 0.020 | 0.017 | 0.353 | 0.335 |

Notes: Standard errors are reported in parentheses. ***, **, and * indicates significance at the 1%, 5%, and 10% level, respectively. [a] the reference group is less than high school. [b] the reference group is less than $30,000 per annum.