PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Variation in the recording of common health conditions in routine hospital data: study using linked survey and administrative data in New South Wales, Australia
AUTHORS	Lujic, Sanja; Watson, Diane; Randall, Deborah; Simpson, Judy; Jorm, Louisa

VERSION 1 - REVIEW

REVIEWER	Lee Nedkoff The University of Western Australia, Australia
REVIEW RETURNED	19-Jun-2014

GENERAL COMMENTS	This study has investigated the level of agreement between self-report and hospital data recording of selected morbidities, and identified the patient and hospital characteristics which contribute to agreement between the data sources. It also used multi-level logistic regression modelling to help explain any variation between hospitals in this regard. This is a well-written manuscript, using sound methodology, and with good interpretation of the salient results. The concordance findings are reasonably consistent with those from other studies, whilst the findings regarding the hospital level factors which influence agreement have implications for funding models. Title: I think this implies that hospitals from across Australia have been included in the study, and given that there may be differences between states in this context, I would amend the title slightly to either include NSW, or to signify that it is an Australian setting/population. Whilst the Methods are generally very well described, clarifying a couple of definitions in the methods would benefit the reader. Methods, Pg 7, line 18: you define the cohort based on those from the 45 and Up study who had an overnight hospitalisation – I would suggest making it clearer that this means (presumably) <u>at least</u> one night in hospital, otherwise it sounds like you are including people with a 1 night stay only. Methods, pg 8, line 10: You have included E15 and E16 in your definition of diabetes? These codes are not usually included, as they refer to nondiabetic coma and other disorders of pancreatic internal

	secretion respectively, so if they were included, this needs explaining. Methods: pg 8, line 10: Because the heart disease and stroke groups are broad disease groups (based on the ICD codes used), a brief outline of the major conditions these definitions include would benefit the reader not familiar with these codes, eg, CHD, dysrhythmias, heart failure etc for heart disease; and noting for the stroke group that it includes codes for cerebrovascular disease without infarction amongst others (which I think is appropriate given that patients will be non-specific in self-reporting stroke). Also, in the footnote to table1, you have presented the heart disease category as I20-25, I26-28, I30-52 – again, for those not familiar with the codes, this implies that some codes have been left out (which they haven't) – it would be less confusing to leave it as I20-I52. Results: One of the limitations of the results in regards to concordance between data sources is that it is presented as an agreement measure by Kappa values only (although the counts for the yes/no combinations for the data sources are provided in Table 2). Therefore we are not told the contribution of false negatives and positives to the level of agreement. I think this should be noted in the Limitations. Page 12, line 8: "Incorporating a 1 year lookback…." This description implies that each patient had a full 1-year lookback period. The general use of this term in the literature is that they have 1 year of hospitalisation history available prior to the 'index' hospital admission. In your study, the 1-year period commences at baseline of the 45 and Up study. Therefore each patient has a varying amount of hospitalisation history available prior to the index hospital admission. I would remove the 'one-year' and just call it a lookback period or previous hospitalisations, as you have done in the headings in Table 2 and Discussion. This means the reader will not assume that a full 1-year of hospitalisation history was available prior to the index admission for all patients. Pg 14, line 28: "The addition of at least one of the four hospital level factors contributed to explaining the residual variation between hospitals for all conditions, except obesity". This statement is a bit confusing as it is very broad, and the hospital level factors which explain the residual variation are different for each morbidity. Discussion, page 15, lines 17-26. It would be worthwhile highlighting here that the lower agreement for heart disease (K=0.47 in your study) compared with other studies may be due to the broad category of heart disease used in your study. Ref 3 found differences in the level of false negatives between CHD, heart failure

	and hypertension, and for example, the coding of MI is usually more accurate in administrative data than other more chronic forms of heart disease. Discussion, page 15, last paragraph. This is a very nice discussion of the issues surrounding the coding of additional diagnoses in Australian administrative hospital data, as these coding standards have major implications for how comorbidities are coded and are often not well understood. The coding of diabetes in particular has undergone many changes over the past decade or so, and deserves mention, because different standards have been applied to diabetes, and therefore it may not be just that it is considered to affect patient management in most hospital stays that it is coded more often. Discussion, pg 18, line 19. You encourage researchers to supplement data with self-report +/- data linkage to overcome the problem of lower concordance in smaller hospitals. A more explicit statement about data linkage in this respect is warranted, because for example, the use of increasing periods of hospitalisation history (beyond the period used in this study) has been shown to increase the accuracy of identifying comorbidities from administrative data, and beyond just stroke and obesity as you have highlighted from your own study in the last sentence of that paragraph. Minor revision: Methods pg 6, line 45 – if you are going to use the full title of ICD, it should read "…Related ***Health*** Problems…"

REVIEWER	Carmen Tsang Lecturer, London School of Hygiene and Tropical Medicine, England
REVIEW RETURNED	21-Jun-2014

GENERAL COMMENTS	Minor point but the Abstract may benefit from moving the content of "Objectives" to the "Outcome measures" section. The "Objectives" would then be replaced by (edited) lines 31-33 of the Introduction, page 5 "This study aimed to further investigate the nature and potential implications of underreporting of morbidity information in administrative hospital data..." This paper was well written and comprehensive in background, reporting and discussion about variation in the recording of comorbidities in hospital records. The use of self-reporting to validate other data sources is relatively novel and this study provides welcome evidence of its use, despite caveats. Importantly, the study highlights the need for triangulation when using administrative health data given inter-hospital variation in recording of the common health conditions investigated. A few minor comments follow: Methods Study population, page 7 line 18 – "…who had an overnight hospitalisation…". Why were day cases not included in the sample?

	Measuring morbidity, page 7 line 47 – "Participants who did not answer the question were excluded…". If data were available, were assessments made of these excluded participants' characteristics compared to those who were included in analyses? Predictors of agreement, page 8 lines 29-38 – Why was ethnicity not measured given variation in self-reporting and in the occurrence of the six morbidities between ethnic groups. Were data on ethnicity not available in the baseline questionnaire? Results Page 12 lines 9-21 – Was there an association between the number of "lookback admissions" and the frequency in which each of the six morbidities were recorded?

REVIEWER	Kim Nazi Department of Veterans Affairs, USA
REVIEW RETURNED	23-Jun-2014

GENERAL COMMENTS	This paper reports a secondary data analysis aimed at investigating the degree to which patient self-report of morbidities varies with administrative hospital data. The topic is interesting and important given that administrative hospital data is often used to enable hospital comparisons with respect to case management and risk adjustments. Performance comparisons of hospitals are increasingly based on this data; thus it is an important inquiry to assess the variance and associated factors. Comments: • Overall the statistical methods are well described. On page 15, the authors note that additional diagnoses are coded only if they affect the treatments received, investigations required, or resources used. Giving this coding 'rule', isn't it true that we would expect to see variance and isn't it really a function of this coding practice to some degree? If this is the case, the findings would need to be clear on this point. Does the inclusion of incorporating previous hospitalizations change the variance and what recommendations do the authors have based on this in order to improvement alignment and decrease variance? Is that also why planned admissions (elective) also have lower odds of having any of the six conditions recorded? This seems to be a direct result of the coding 'rules'. • In reading the paper, the distinction between use of medical records and patient self-report emerges, however I think this point could be made earlier in the paper since it is likely that most readers will find this to be an important distinction. It will be important to describe the differences between medical records and administrative hospital data (e.g., in general how medical record review and coding generates administrative data). Are there any indications that lack of documentation of these comorbidities in the administrative hospital data may also be lack of documentation in the medical record? If so, this is an equally compelling finding. Without understanding this distinction, it is hard to fully understand the implications of the findings. This first arises in the abstract in the objectives—to investigate the agreement between [patient] self-report and [clinician] recording? The reader needs to fully understand how administrative hospital data is generated (isn't it based on coding of the medical record?) to understand the root causes of the variance. • Depth of coding emerges as a factor which has significant impact but is never fully defined. On page 17 this is also characterized as crude: "Even though the measure of depth of coding we used was

	crude, and related to hospital size, it still helps in highlighting the impact of coding practices on variations among hospitals." It will be important to define this factor and explicitly state how it was measure in this analysis. It may need to be listed as a limitation. • I think it would strengthen the paper to clearly address these questions up front. Perhaps clearly framing the paper as quantifying the known variance that results from these factors. • Study limitations should also reference the 18% participation rate in terms of generalizability. • Minor comments o Page 5 paragraph 2, suggest using the more quantitative term "greater" underreporting rather than the potentially misinterpreted "worse" underreporting o Table 1 could be split into two tables and show the Hospital characteristics as a separate table, avoiding a blank "all hospitals" column for demographics

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 – Lee Nedkoff

1. Title: I think this implies that hospitals from across Australia have been included in the study, and given that there may be differences between states in this context, I would amend the title slightly to either include NSW, or to signify that it is an Australian setting/population.

Response: Thank you for the suggestion, we have amended the title to "Variation in the recording of common health conditions in routine hospital data: study using linked survey and administrative data in New South Wales, Australia".

2. Whilst the Methods are generally very well described, clarifying a couple of definitions in the methods would benefit the reader. Methods, Pg 7, line 18: you define the cohort based on those from the 45 and Up study who had an overnight hospitalisation – I would suggest making it clearer that this means (presumably) at least one night in hospital, otherwise it sounds like you are including people with a 1 night stay only.

Response: We have amended the text on page 8 to read: "The study population comprised patients aged 45 years and above who participated in the 45 and Up Study and who had a hospitalization lasting at least one night in the period up to 365 days prior to filling out the baseline 45 and Up Study survey."

3. Methods, pg 8, line 10: You have included E15 and E16 in your definition of diabetes? These codes are not usually included, as they refer to nondiabetic coma and other disorders of pancreatic internal secretion respectively, so if they were included, this needs explaining.

Response: The reviewer is correct; diabetes should be coded as E10-E14. The reference in the manuscript to E10-E16 was incorrect, and a typographical error. The ICD-10-AM codes to produce the results actually were E10-E14. This change has been made to the methods section (page 9) of the manuscript only, the reference in the table footnotes was correct.

4. Methods: pg 8, line 10: Because the heart disease and stroke groups are broad disease groups

(based on the ICD codes used), a brief outline of the major conditions these definitions include would benefit the reader not familiar with these codes, eg, CHD, dysrhythmias, heart failure etc for heart disease; and noting for the stroke group that it includes codes for cerebrovascular disease without infarction amongst others (which I think is appropriate given that patients will be non-specific in self-reporting stroke). Also, in the footnote to table1, you have presented the heart disease category as I20-25, I26-28, I30-52 – again, for those not familiar with the codes, this implies that some codes have been left out (which they haven't) – it would be less confusing to leave it as I20-I52.

Response: Thank you for the suggestion. We have amended the footnote to Table1 to state "I20-I52". We have also added a sentence to the methods section (page 9): "The inclusion of broader ICD-10-AM codes for heart disease and stroke was chosen because of the broad definition of disease type in the self-reported data. Thus, heart disease codes were inclusive of coronary heart disease, pulmonary heart disease, and other forms of heart diseases including heart failure and arrhythmias. Stroke codes included cerebrovascular diseases without infarction among others."

5. Results:
One of the limitations of the results in regards to concordance between data sources is that it is presented as an agreement measure by Kappa values only (although the counts for the yes/no combinations for the data sources are provided in Table 2). Therefore we are not told the contribution of false negatives and positives to the level of agreement. I think this should be noted in the Limitations.

Response: We have chosen to present the results as Kappa values only because neither hospital nor self-report data could be considered a 'gold standard'. However Table 2 contains complete information about the numbers of false positives and false negatives, for interested readers.

6. Page 12, line 8: "Incorporating a 1 year lookback…." This description implies that each patient had a full 1-year lookback period. The general use of this term in the literature is that they have 1 year of hospitalisation history available prior to the 'index' hospital admission. In your study, the 1-year period commences at baseline of the 45 and Up study. Therefore each patient has a varying amount of hospitalisation history available prior to the index hospital admission. I would remove the 'one-year' and just call it a lookback period or previous hospitalisations, as you have done in the headings in Table 2 and Discussion. This means the reader will not assume that a full 1-year of hospitalisation history was available prior to the index admission for all patients.

Response: Although each patient had varying amounts of hospitalisation history, each patient in the present study did have a full 1-year lookback period, because the hospital data collection spanned 2000 onwards and the baseline 45 and Up survey dates ranged from January 2006 to December 2009. As the dates of admission and survey completion were available for all participants, we were able to determine a date of the 'index admission' (closest admission date to the survey completion, but no longer than 365 days prior), as well as all the admissions in the full year preceding the 'index admission'.

7. Pg 14, line 28: "The addition of at least one of the four hospital level factors contributed to explaining the residual variation between hospitals for all conditions, except obesity".
This statement is a bit confusing as it is very broad, and the hospital level factors which explain the residual variation are different for each morbidity.

Response: The sentence describes our finding that hospital factors (type, size, remoteness, depth of coding) help explain residual variation in the outcome for 5 out of 6 studied conditions. The amount of variation attributable to each factor, as the reviewer rightfully pointed out, is different for each morbidity and is summarised in the sentence that follows.

To avoid ambiguity, we have re-worded the sentence on page 16 to read “The addition of hospital-level factors contributed to explaining (i.e. decreasing) the residual variation for all conditions, except obesity.”

8. Discussion, page 15, lines 17-26. It would be worthwhile highlighting here that the lower agreement for heart disease (K=0.47 in your study) compared with other studies may be due to the broad category of heart disease used in your study. Ref 3 found differences in the level of false negatives between CHD, heart failure and hypertension, and for example, the coding of MI is usually more accurate in administrative data than other more chronic forms of heart disease.

Response: Thank you for the suggestion. We have inserted the following sentence on page 16: “Lower agreement rates for heart disease could be due to the broader range of heart disease types included in our study, with known lower levels of agreement for heart failure compared to myocardial infarction.[9, 40]”.

9. Discussion, page 15, last paragraph. This is a very nice discussion of the issues surrounding the coding of additional diagnoses in Australian administrative hospital data, as these coding standards have major implications for how comorbidities are coded and are often not well understood. The coding of diabetes in particular has undergone many changes over the past decade or so, and deserves mention, because different standards have been applied to diabetes, and therefore it may not be just that it is considered to affect patient management in most hospital stays that it is coded more often.

Response: We have added a sentence and a new reference in response to the reviewer’s comments on page 17 as follows: “On the other hand, we found that diabetes is well recorded, suggesting that it is considered to affect patient management in most hospital stays, and possibly reflecting the impact of changes to the Australian Coding Standards for diabetes such that between 2008 and 2010 diabetes with complications could be coded even where there was no established cause and effect relationship between diabetes and the complication.[43]”.

10. Discussion, pg 18, line 19. You encourage researchers to supplement data with self-report +/- data linkage to overcome the problem of lower concordance in smaller hospitals. A more explicit statement about data linkage in this respect is warranted, because for example, the use of increasing periods of hospitalisation history (beyond the period used in this study) has been shown to increase the accuracy of identifying comorbidities from administrative data, and beyond just stroke and obesity as you have highlighted from your own study in the last sentence of that paragraph.

Response: We agree with the reviewer, and are in favour of using data linkage, where possible. In situations where data linkage is not an option or past hospitalisations unavailable, supplementing the information with at least self-report is warranted. We have included the text and a new reference in the Discussion (page 17) to reflect the benefit of incorporating previous hospitalisations:

“It is for these reasons that researchers using administrative datasets are encouraged to incorporate information from previous hospitalizations, to increase the likelihood of capturing morbidity, as demonstrated in this as well as other Australian studies. [44]”

11. Minor revision:
Methods pg 6, line 45 – if you are going to use the full title of ICD, it should read “…Related Health Problems…”

Response: The text is amended to read “International Statistical Classification and Related Health Problems”.

12. I believe the reference list probably needs some formatting attended to.

Response: The reference list has been re-formatted. New references have also been added in response to reviewers’ comments:

23. Eagar K. Counting acute inpatient care. ABF Information Series No 5 2010. https://ahsri.uow.edu.au/content/groups/public/@web/@chsd/documents/doc/uow082637.pdf (accessed July 2014).
43. Knight L, Halech R, Martin C, et al. Impact of changes in diabetes coding on Queensland hospital principal diagnosis morbidity data. 2011. http://www.health.qld.gov.au/hsu/tech_report/techreport_9.pdf (accessed July 2014).
44. Preen DB, Holman CDAJ, Spilsbury K, et al. Length of comorbidity lookback period affected regression model performance of administrative health data. J Clin Epidemiol 2006;59(9):940-46

Reviewer 2: Carmen Tsang

1. Minor point but the Abstract may benefit from moving the content of "Objectives" to the "Outcome measures" section. The "Objectives" would then be replaced by (edited) lines 31-33 of the Introduction, page 5 "This study aimed to further investigate the nature and potential implications of underreporting of morbidity information in administrative hospital data...".

Response: We have amended the abstract to read:

“Objective: To investigate the nature and potential implications of underreporting of morbidity information in administrative hospital data”.
“Outcome measures: Agreement between self-report and recording of six morbidities in administrative hospital data, and between-hospital variation and predictors of positive agreement between the two data sources.”

2. A few minor comments follow: Methods
Study population, page 7 line 18 – “…who had an overnight hospitalisation…”. Why were day cases not included in the sample?

Response: Day stay patients were excluded from the analysis to make our results more robust and generalisable beyond NSW and Australia. Australia is one of only a few countries that formally ‘admits’ same day patients as ‘inpatients’, while in most other comparable countries, including the USA, Canada and the UK, an ‘inpatient’ is a patient who stays in hospital for at least one night (Eagar 2010). Additionally, even within Australia, there is considerable variation among jurisdictions in whether or not certain services such as dialysis and chemotherapy are classified as admitted patient day-stays or outpatient services.

We have added a sentence and a new reference in response to the reviewer’s comments on page 8 as follows:
“Day stay patients were excluded from the analysis to make the study more robust and generalizable

beyond NSW and Australia, as there are differences in admission practices for day stay patients between Australia and most other comparable countries. [23]".

Eagar K. Counting acute inpatient care. ABF Information Series No 5 2010. https://ahsri.uow.edu.au/content/groups/public/@web/@chsd/documents/doc/uow082637.pdf (accessed July 2014).

3. Measuring morbidity, page 7 line 47 – "Participants who did not answer the question were excluded…". If data were available, were assessments made of these excluded participants' characteristics compared to those who were included in analyses?

Response: We did not include information on the differences between participants who were and weren't included in the analysis in the paper, because only a small proportion (3.65%) of participants did not answer the question about morbidities, and thus any imputation of the missing data would have had negligible impact on the results. However, we've produced a table below for the information of reviewers. Slightly higher proportions of participants excluded from the analysis were: female, overseas born, and with lower educational attainment. Those excluded also had significantly higher proportions of missing information about education, household income and functional capacity. In terms of morbidities recorded in hospital data, diabetes and hypertension were more prevalent among those in the study, and smoking was more prevalent among those that were excluded. All the statistical models in the study included an adjustment for demographic factors.

Characteristics	In the study (N = 32,832)		Not in the study (N = 1,242)	
	N	%	N	%
Sex				
Male	16,812	51.2%	552	44.4%
Female	16,020	48.8%	690	55.6%
Age				
45-59	9,666	29.4%	433	34.9%
60-79	16,624	50.6%	527	42.4%
80+	6,540	19.9%	280	22.5%
Country of birth				
Australia	25,001	76.1%	871	70.1%
Other	7,448	22.7%	325	26.2%
Unknown	383	1.2%	46	3.7%
Highest education level				
No school	5,196	15.8%	234	18.8%
Year 10 or equivalent	7,894	24.0%	300	24.2%
Year 12 or equivalent	2,975	9.1%	144	11.6%
Trade	4,270	13.0%	155	12.5%
Certificate	6,109	18.6%	171	13.8%
University degree	5,662	17.2%	166	13.4%
Unknown	726	2.2%	72	5.8%
Household income ($, per annum)				
<20,000	9,077	27.6%	191	15.4%
20,000 - <50,000	8,223	25.0%	160	12.9%
50,0000 - <70,000	2,560	7.8%	53	4.3%
70,000+	5,042	15.4%	78	6.3%
Not disclosed	6,003	18.3%	130	10.5%
Missing	1,927	5.9%	630	50.7%
Functional status				

No limitation 4,915 15.0% 156 12.6%
Mild limitation 6,011 18.3% 102 8.2%
Moderate limitation 8,701 26.5% 110 8.9%
Severe limitation 10,121 30.8% 163 13.1%
Missing 3,084 9.4% 711 57.2%
Morbidities (from hospital data)
Hypertension 6,201 18.9% 190 15.3%
Heart disease 5,581 17.0% 208 16.8%
Diabetes 3,907 11.9% 100 8.05%
Stroke 847 2.58% 32 2.58%
Smoking 1,932 5.9% 104 8.4%
Obesity 665 2.0% 20 1.6%

4. Predictors of agreement, page 8 lines 29-38 – Why was ethnicity not measured given variation in self-reporting and in the occurrence of the six morbidities between ethnic groups. Were data on ethnicity not available in the baseline questionnaire?

Response: The baseline questionnaire included questions on country of birth and ancestry. Our study did include an adjustment for country of birth in all the multilevel models, alongside other demographic factors. As 76% of the participants were Australian born, the country of birth variable was analysed as a dichotomous one (Australian-born, overseas-born). Supplementary Table 2 gives odds ratios and 95% confidence intervals for the country of birth variable. The results show that there is no statistically significant difference in the odds of recording of morbidities between those born in Australia versus other countries.

5. Results
Page 12 lines 9-21 – Was there an association between the number of "lookback admissions" and the frequency in which each of the six morbidities were recorded?

Response: A priori, one would expect positive correlation between the number of lookback admissions and the frequency of recording of morbidities, because there is an increased likelihood of capturing morbidity with an increase in the number of previous admissions screened. For our study, the correlation between the number of previous admissions and number of instances each of the morbidities was recorded in hospital data is:

Heart disease Diabetes Stroke Hypertension Smoking Obesity
Correlation* = 0.616 0.824 0.287 0.536 0.782 0.253
* correlation among those that have identified morbidity using a self-report

The value of incorporating lookback periods to help ascertain morbidity has been reported in studies such as Preen at al 2006. We have chosen a period of 1-year lookback for our study, because it was found to be sufficient to identify morbidity for both medical and surgical patients (Preen et al 2006). The examination of varying degrees of lookback period was outside the scope of this study.
Preen DB, Holman CDJ, Spilsbury K, et al. Length of comorbidity lookback period affected regression model performance of administrative health data. J Clin Epidemiol 2006; 59: 940-6.

Reviewer 3: Kim Nazi

1. On page 15, the authors note that additional diagnoses are coded only if they affect the treatments received, investigations required, or resources used. Giving this coding 'rule', isn't it true that we would expect to see variance and isn't it really a function of this coding practice to some degree? If

this is the case, the findings would need to be clear on this point. Does the inclusion of incorporating previous hospitalizations change the variance and what recommendations do the authors have based on this in order to improvement alignment and decrease variance? Is that also why planned admissions (elective) also have lower odds of having any of the six conditions recorded? This seems to be a direct result of the coding 'rules'.

Response: The coding rule, which governs coding practice in Australia, impacts on whether a clinical coder codes a diagnosis for a particular stay in hospital, despite the fact that patient's medical records may contain such information. Thus, as a direct consequence of the 'coding rule', one would anticipate underreporting of morbidities in administrative datasets in Australia (as was indeed shown to be the case in a study by Powell et al 2001) but one would not necessarily expect between-hospital variation. Hence, our study aimed to quantify whether there are systematic differences in recording of health conditions between hospitals and to identify factors that explain these differences.

Regarding whether the inclusion of previous hospitalisations changes the variance – yes, it does – the incorporation of previous hospitalisations reduced the variance, as noted in the original manuscript on page 14, lines 10-12. This conclusion was based on the results presented in Figure 1.

Powell H, Lim LL, Heller RF. Accuracy of administrative data to assess comorbidity in patients with heart disease: an Australian perspective. J Clin Epidemiol 2001;54(7):687-93

2. In reading the paper, the distinction between use of medical records and patient self-report emerges, however I think this point could be made earlier in the paper since it is likely that most readers will find this to be an important distinction. It will be important to describe the differences between medical records and administrative hospital data (e.g., in general how medical record review and coding generates administrative data). Are there any indications that lack of documentation of these comorbidities in the administrative hospital data may also be lack of documentation in the medical record? If so, this is an equally compelling finding. Without understanding this distinction, it is hard to fully understand the implications of the findings. This first arises in the abstract in the objectives—to investigate the agreement between [patient] self-report and [clinician] recording? The reader needs to fully understand how administrative hospital data is generated (isn't it based on coding of the medical record?) to understand the root causes of the variance.

Response: We have clarified the distinction between administrative data and medical records and modified the manuscript to include further information about how administrative data are generated. In the Discussion section, we note that the discrepancies in the levels of coding could be driven by coding rules, as suggested by the reviewer. We also discuss the factors that might explain the variation in the depth of coding (please see page 17, lines 37-51 of the original manuscript). The relative roles of lack of documentation in the medical record and of coding rules in contributing to underreporting in administrative data are difficult to investigate without a detailed review of medical records, which was beyond the scope of this study.

The following changes have been made to the manuscript, in order to increase clarity:

Introduction, page 5:
"Most methods of case-mix adjustment rely principally on demographic and diagnostic information that is captured in administrative hospital data collections. The hospital data is collected and recorded in a data base for administrative purposes, with clinical coders coding diagnostic information based on the patient's medical records. [3]"

Methods, page 8:
“Additional diagnoses are defined as ‘a condition or complaint either coexisting with the principal diagnosis or arising during the episode of care’ in the Australian Coding Standards and should be interpreted as conditions that affect patient management. [21] Assignment of diagnosis codes is done by trained clinical coders, using information from the patient’s medical records.”

3. Depth of coding emerges as a factor which has significant impact but is never fully defined. On page 17 this is also characterized as crude: “Even though the measure of depth of coding we used was crude, and related to hospital size, it still helps in highlighting the impact of coding practices on variations among hospitals.” It will be important to define this factor and explicitly state how it was measure in this analysis. It may need to be listed as a limitation.

Response: The definition of depth of coding we used is presented on page 8 of the original manuscript (lines 51-57).To clarify the definition, we have now amended the text to read:

“Depth of hospital coding was the mean number of additional diagnoses coded per episode of care for each hospital, calculated using all overnight hospitalizations for the full 45 and Up Study cohort from 2000 to 2010, and divided into four groups at the 25th, 50th and 75th percentile”.

We have also included a sentence in the methods (page 7) to refer to episodes of care as:
“The APDC includes records of all public and private hospital admissions ending in a separation, i.e. discharge, transfer, type-change or death. Each separation is referred to as an episode of care.”

4. I think it would strengthen the paper to clearly address these questions up front. Perhaps clearly framing the paper as quantifying the known variance that results from these factors.

Response: We have addressed this comment in our reframing of the research question in response to the Editorial Comment 1.

5. Study limitations should also reference the 18% participation rate in terms of generalizability.

Response: Even though the response rate was 18%, the 45 and Up Study includes 1 in 10 persons aged 45 and over in New South Wales. The 45 and Up Study sample has excellent heterogeneity, and participants resemble the general population in many ways, albeit they are in general of higher socioeconomic status and more ‘healthy’ (Mealing et al). Relative measures of effect (odds ratios) have been shown not to be biased. We have described the potential generalisability limitation in the Discussion section by adding the following text on page 21:

“Although the 45 and Up Study had a response rate of 18%, the study sample is very large and has excellent heterogeneity. Furthermore, exposure-outcome relationships estimated from the 45 and Up Study data have been shown to be consistent with a large ‘representative’ population survey of the same population.[55]”

6. Minor comments
Page 5 paragraph 2, suggest using the more quantitative term “greater” underreporting rather than the potentially misinterpreted “worse” underreporting

Response: We have replaced the term “worse” with term “greater” as suggested.

7. Table 1 could be split into two tables and show the Hospital characteristics as a separate table, avoiding a blank “all hospitals” column for demographics

Response: We have split Table 1 into two tables, and modified the references throughout the Results section accordingly.

VERSION 2 – REVIEW

REVIEWER	Lee Nedkoff The University of Western Australia, Australia
REVIEW RETURNED	29-Jul-2014

GENERAL COMMENTS	My comments and queries have all been fully and well addressed by the authors responses, and I therefore don't have any further comments to make regarding this manuscript. As suggested, a statistical review would be useful as I am not qualified to comment in detail on the some of the statistical methods used in this study. I have suggested 'Accept' below as my recommendation, pending statistical review.

REVIEWER	Kim M. Nazi Department of Veterans Affairs USA
REVIEW RETURNED	18-Jul-2014

GENERAL COMMENTS	This version of the paper is significantly stengthened by the revisions made in response to peer reviews. On page 14 suggest chaning "one at the time" to "one at a time".