

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A cohort study on physician documentation and the accuracy of administrative data coding to improve passive surveillance of transient ischemic attacks
AUTHORS	Yu, Amy Ying Xin; Quan, Hude; McRae, Andrew; Wagner, Gabrielle; Hill, Michael; Coutts, Shelagh B.

VERSION 1 - REVIEW

REVIEWER	Ruth Hall Institute for Clinical Evaluative Sciences, Canada
REVIEW RETURNED	11-Jan-2017

GENERAL COMMENTS	Important work but needs to be more focused and concise.
-------------------------	----------------------------------------------------------

REVIEWER	Jonathan Thigpen Notre Dame of Maryland University School of Pharmacy United States
REVIEW RETURNED	18-Jan-2017

GENERAL COMMENTS	<p>This paper provides important evidence regarding the accuracy of TIA codes and more importantly, what factors are associated with coding accuracy. With increasing utilization of administrative data in research and clinical practice, it is essential to better understand the accuracy and utility of these codes.</p> <p>Overall, this paper is pertinent and adds to our current understanding of this topic. It also provides unique insight into the role of physician documentation and resulting coding accuracy. I especially liked Table 1. It was very thoughtful to include and complemented the paper greatly.</p> <p>However, in a paper such as this, the methods/results of re-abstractation and accuracy among the different coding algorithms is difficult to follow at times. Accordingly, the tables/figures and the text need to align extremely well and the text needs to clearly convey important details. Unfortunately, although infrequently, I was stuck trying to figure out where some of the numbers were coming from (some of the presented results and discussion were not directly included in the tables/figures) or trying to interpret what patient cohort the authors were referring to at a given time (terminology used is not consistent at times). Accordingly, reading the paper was at times a little jarring and confusing. Overall the authors did a very good job and addressing the minor edits/comments below will help lead to better readability.</p> <p>Thank you for your time and effort in completing this pertinent project; which will ultimately aid many future clinicians and researchers.</p>
-------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Please address my comments below.

1. Abstract; Results; Page 2

“The majority of disagreement (n=103, 84.4%) arose from re-abstracted TIA cases that were misclassified in administrative data coding.”

Comment: For clarification, please specify that these cases were misclassified as non-TIA (in the main position) when in fact they were TIA. Please edit the sentence to something such as “...misclassified as non-TIA in administrative data coding.”

2. Introduction; Page 3; Paragraph 2

Comment: The literature also indicates that coding accuracy is largely determined by code position (primary vs. non-primary), clinical setting (inpatient vs. outpatient), date of code use (e.g. 1970 vs. 2010), type of code (ICD 9 vs. ICD 10), and disease of interest (certain diseases have higher PPVs than others). Authors should consider adding this information.

3. General Comment

Comment: Throughout the paper, consider changing the term “re-abstractation” to an easier to follow term such as “manual review” or maybe “adjudication”.

4. Results; Page 8; Paragraph 2

“Inter-abstractor agreement between the two stroke neurologists was 76.7% ($\kappa = 0.50$) for the diagnosis of TIA.”

Comment: Please address this in the strengths and limitations section of the discussion. Is this appropriate? What led to this? Could this have influenced the results?

Comment: In the methods, please clarify how cases were resolved when the reviewers disagreed and had differing final verdicts. Were there additional adjudication steps? How was the final verdict determined? I ask because there seemed to be a good amount of disagreements in cases (76.7% agreement).

5. Results; Page 8; Paragraph 2

“The main-position NACRS algorithm identified 79 cases of TIA and the any-position algorithm identified 151 cases.”

Comment: As it reads, this statement is confusing. Elsewhere, “cases” are in reference to the true TIAs deemed so after manual re-abstractation (n=163). “Cases” are also sometimes in reference to the misclassified “cases”. However, the “cases” in this sentence refer to the instances in which they were coded for TIA and re-abstractation confirmed or denied (n=60 confirmed and n=19 denied). Does this make sense? For clarification, the statement should be changed to something like....“The NACRS algorithm identified 79 patients coded with TIA in the main position and 151 coded in any-position.”

Also, to promote better flow, this statement should be placed after

	<p>the sentence “Figure 1 shows the distribution of cases of disagreement between chart re-abstraction and the two TIA coding algorithms.”</p> <p>6. Results; Page 9; Paragraph 1</p> <p>“The relative risk of disagreement between chart re-abstraction and administrative data coding when the final diagnosis was uncertain or absent was 1.82 [CI95 1.36 , 2.44]. The risk difference was 18.5%.”</p> <p>Comment: This information doesn’t seem to be located in any of the tables. To clarify, was this a combined RR for “uncertain” or “absent”? If so, please edit the statement to something like..... “As a combined measure, the relative risk of disagreement....” This seems necessary since the tables/figures report them as separate measures (uncertainty vs. lack of definitive diagnosis), yet the text here provides them as a combined measure.</p> <p>7. Discussion; Page 10; Paragraph 2</p> <p>“Physician documentation was the dominant factor influencing the accuracy of the administrative data coding for TIA and minor stroke patients discharged from the ED. The risk difference was 18.5%, suggesting that close to one in five charts with coding misclassifications could have improved accuracy if a definitive diagnosis was documented.”</p> <p>Comment: Since this is a combined measure (risk difference 18.5%), please change this statement to something like “... suggesting that coding misclassifications could have been prevented in approximately one in every five charts if a definitive diagnosis was documented or uncertainty removed.”</p> <p>8. Results/Discussion</p> <p>Comment: There is no analysis/discussion regarding comparing accuracy as a whole for the “main position” vs. the “any position” algorithms. This is not your intended purpose of this study, so I don’t expect you to perform additional analyses. However, it is very interesting that your re-abstracted TIA cases have higher coding accuracy in “any position” (63.8%) vs. the “main position” (36.8%). I have several theories as to why this might have occurred, but I’d like you to briefly comment on this in the discussion section. I think this is beneficial to discuss since a vast amount of research indicates that codes in the primary position tend to have higher positive predictive values compared to codes in a non-primary position. Yet your results refute this in this instance.</p>
--	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

REVIEWER	Eric Adelman, MD University of Michigan Ann Arbor, MI USA
REVIEW RETURNED	24-Jan-2017

GENERAL COMMENTS	The authors aimed to evaluate the accuracy of TIA coding in the ED and examined predictors of accurate coding. The patient data used in these analyses came from 2 TIA trials done at a tertiary care hospital. Charts were abstracted by a vascular neurologist and a
-------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

second independent review was performed on a subset of charts. TIA was defined by duration of symptoms rather than absence of brain infarction. The authors found that 39% of patients the stroke neurologists thought had TIAs were not coded at TIAs. Diagnostic uncertainty predicted disagreement between the stroke neurologist review and coder. Ongoing symptoms while the patient was in the ED predicted concordance between the vascular neurology review and coding.

This is an interesting study as administrative data is frequently used in epidemiologic studies and for quality reporting. The methods are appropriate but would have been strengthened by having a coder review the charts in conjunction with the stroke neurologists. Additionally, selection bias is a major limitation of this work. By only including patients enrolled in a TIA trial, the authors enriched their sample with patients at higher likelihood of having a TIA. As the authors note, this lessens the generalizability of the study. There is also a tension between diagnostic uncertainty at the time of presentation and how that uncertainty is documented. The authors seem to presume that a definite diagnosis of TIA can always be made in the ED. Perhaps it is easier to make a TIA diagnosis in retrospect and this would impact the authors' conclusions. The recommendations about improving documentation when there is diagnostic uncertainty are well said.

Specific comments:

Strengths and limitations section

--Line 6: Consider reframing the first bullet point. If the administrative data is not good at identifying TIAs in the ED, data accuracy is not particularly relevant. Perhaps putting the emphasis on predictors of data accuracy as a way to improve the sensitivity of administrative data would emphasize this point.

--Line 20: The authors refer to the "patient-physician" interaction, but it is not clear what this means.

--Consider noting that this study was done in patients participating in a TIA trial. That seems to be a bigger limitation than the fact it was done at a university teaching hospital.

Introduction

--Page 4, line 10: These are patients with TIAs discharged from the ED, not TIAs discharged from the ED.

--Page 4, line 19: It is not clear what the authors mean by "clinical acumen"

--Page 4, line 29: If the authors are going to highlight physician factors, be specific about what factors those are. Reading this, one thinks of role (trainee, APP, attending) or time in practice, as physician factors but the authors do not have this information.

Methods

--Page 5, line 25: The authors should highlight why they used a time based definition of TIA, rather than the contemporary tissue based definition (Stroke 2009;40:2276-2293)?

--In the methods section the authors may want to comment that they were unable to adjust for ED and neurology provider characteristics, specifically role (attending, APP, resident, stroke specialists, etc) and experience (e.g. years in practice.)

Discussion

--Page 10, line 31: It seems to be valid that ongoing symptoms and

	<p>neurology consultation improve coding, but it is not clear how male sex impacts coding accuracy.</p> <p>--How much knowledge do physicians have of coding? The authors talk obliquely about this, but their work would be strengthened by addressing this issue more directly. If physicians aren't knowledgeable about coding requirements then it is not surprising that these requirements are not met. In a similar vein, more concrete recommendations on how to improve coding would be useful.</p> <p>Conclusions</p> <p>--The manuscript would be strengthened if the authors were more specific about their findings in this section.</p> <p>Tables</p> <p>Table 2: consider adding another column with the results of an appropriate statistical test of difference between agreement and disagreement.</p>
--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

REVIEWER	Cheng-Yang Hsieh Department of Neurology, Tainan Sin Lau Hospital, TAIWAN
REVIEW RETURNED	02-Feb-2017

GENERAL COMMENTS	<p>This study aims to explore the relationship between physician documentation and the accuracy of administrative data coding in TIA patients. In general, the manuscript is well-written. And I totally agreed that physician documentation is very important for coding accuracy in paroxysmal disorders such as TIA. I have the following comments for the authors in the future revision:</p> <ol style="list-style-type: none"> 1. Title: Is this really a COHORT study? Why? 2. Line 54 of Page 4: "Both studies require at least one clinical evaluation by a neurologist". But in Table 2, 75.4%-84.4% of patients had neurology consult, not 100%. 3. Line 26 of Page 5: The authors adopt the WHO time-based criteria for diagnosing TIA when chart re-abstraction. For a patient with symptoms lasting less than 24 hours but a MRI showing evidence of tissue ischemia, the diagnosis would be adjudicated as a TIA, but not a stroke. Will the different definition of TIA (WHO time-based vs. MRI-based) affect results of the present study? Will the administrative coders give the ICD-10 code of cerebral infarct just based on the MRI results? 4. Line 31 of Page 5: A senior stroke neurologist independently reviewed a random sample of thirty charts to determine inter-abstractor reliability. And in Line 45 of Page 5, the agreement was only modest (76.7%, kappa = 0.50). How to explain this disagreement between two stroke neurologists? Given such a modest agreement of TIA diagnosis between stroke neurologists, it may not be surprising for low agreement between physician document and administrative coders in diagnosis of TIA. 5. Line 45 of Page 6: "a 85-year-old woman..." It should be a typo ("82F" in Table 1). 6. Discussion: Will other factors, such as insurance's policy for reimbursement, affect the coding behavior? For example, if the reimbursement ED visits for ischemic stroke is much higher than ED visits for TIA, will the coder more likely to code ischemic stroke?
-------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 Reviewer Name: Ruth Hall
Institution and Country: Institute for Clinical Evaluative Sciences, Canada
Competing Interests: None declared
Important work but needs to be more focused and concise.

Authors response: Thank you for your review and interest. We trust that our message is more focused with the revisions, highlighted in tracked change.

Reviewer: 2 Reviewer Name: Jonathan Thigpen
Institution and Country: Notre Dame of Maryland University School of Pharmacy, United States
Competing Interests: None declared

This paper provides important evidence regarding the accuracy of TIA codes and more importantly, what factors are associated with coding accuracy. With increasing utilization of administrative data in research and clinical practice, it is essential to better understand the accuracy and utility of these codes.

Overall, this paper is pertinent and adds to our current understanding of this topic. It also provides unique insight into the role of physician documentation and resulting coding accuracy. I especially liked Table 1. It was very thoughtful to include and complemented the paper greatly.

However, in a paper such as this, the methods/results of re-abstraction and accuracy among the different coding algorithms is difficult to follow at times. Accordingly, the tables/figures and the text need to align extremely well and the text needs to clearly convey important details. Unfortunately, although infrequently, I was stuck trying to figure out where some of the numbers were coming from (some of the presented results and discussion were not directly included in the tables/figures) or trying to interpret what patient cohort the authors were referring to at a given time (terminology used is not consistent at times). Accordingly, reading the paper was at times a little jarring and confusing. Overall the authors did a very good job and addressing the minor edits/comments below will help lead to better readability.

Thank you for your time and effort in completing this pertinent project; which will ultimately aid many future clinicians and researchers.

Please address my comments below.

1. Abstract; Results; Page 2

“The majority of disagreement (n=103, 84.4%) arose from re-abstracted TIA cases that were misclassified in administrative data coding.”

Comment: For clarification, please specify that these cases were misclassified as non-TIA (in the main position) when in fact they were TIA. Please edit the sentence to something such as “...misclassified as non-TIA in administrative data coding.”

Authors response: Thank you for your interest and thoughtful comments. We have made the recommended clarification in the Abstract.

2. Introduction; Page 3; Paragraph 2

Comment: The literature also indicates that coding accuracy is largely determined by code position (primary vs. non-primary), clinical setting (inpatient vs. outpatient), date of code use (e.g. 1970 vs. 2010), type of code (ICD 9 vs. ICD 10), and disease of interest (certain diseases have higher PPVs than others). Authors should consider adding this information.

Authors response: We agree this is important background information and added the following, with references (page 4, end of first paragraph):

“Additionally, the accuracy of administrative data codes is dependent on the disease under study, the study time period, the International Classification of Diseases (ICD) iteration used⁶, the case definition algorithm (choice of ICD codes and their positions)^{7 8}, and the clinical setting^{8 9}.

3. General Comment

Comment: Throughout the paper, consider changing the term “re-abstraction” to an easier to follow term such as “manual review” or maybe “adjudication”.

Authors response: Thank you, we replaced “re-abstraction” with the suggested terms.

4. Results; Page 8; Paragraph 2

“Inter-abstractor agreement between the two stroke neurologists was 76.7% ($\kappa = 0.50$) for the diagnosis of TIA.”

Comment: Please address this in the strengths and limitations section of the discussion. Is this appropriate? What led to this? Could this have influenced the results?

Authors response: While it is true that the agreement reported was moderate, this finding is consistent with the TIA literature. The diagnosis of TIA is clinical, often without objective confirmatory tests. In the discussions, we added a new paragraph with appropriate references (page 13, second paragraph):

“In the subset of charts that underwent a second independent review, we showed that the agreement was 76.7%. This is consistent with a study conducted in the Stanford TIA clinic that reported a 72% overall agreement among fellowship-trained stroke neurologists.²⁸ The subjectivity in TIA diagnosis, especially during the initial health encounter, is impossible to eliminate. Using a tissue-based diagnosis for TIA²⁹ based on an objective biologic endpoint has potential to improve the inter-rater reliability. However, cerebral ischemia can only be demonstrated in 30-70% of clinically-diagnosed TIA using magnetic resonance diffusion-weighted imaging and the finding of ischemia is dependent on symptom duration.³⁰ Further, because diffusion-weighted imaging is not routinely utilized in clinical practice, applying a tissue-based TIA diagnosis as reference standard would negatively impact the generalizability of our results.”

Comment: In the methods, please clarify how cases were resolved when the reviewers disagreed and had differing final verdicts. Were there additional adjudication steps? How was the final verdict determined? I ask because there seemed to be a good amount of disagreements in cases (76.7% agreement).

Authors response: The senior stroke neurologist (SBC) independently reviewed a random sample of thirty charts. We did not have two independent adjudications of all charts. Based on the literature, we predetermined that more charts will be independently adjudicated if the agreement was less than 70%.

5. Results; Page 8; Paragraph 2

“The main-position NACRS algorithm identified 79 cases of TIA and the any-position algorithm identified 151 cases.”

Comment: As it reads, this statement is confusing. Elsewhere, “cases” are in reference to the true TIAs deemed so after manual re-abstraction (n=163). “Cases” are also sometimes in reference to the misclassified “cases”. However, the “cases” in this sentence refer to the instances in which they were coded for TIA and re-abstraction confirmed or denied (n=60 confirmed and n=19 denied). Does this make sense? For clarification, the statement should be changed to something like....“The NACRS algorithm identified 79 patients coded with TIA in the main position and 151 coded in any-position.” Also, to promote better flow, this statement should be placed after the sentence “Figure 1 shows the distribution of cases of disagreement between chart re-abstraction and the two TIA coding algorithms.”

Authors response: Thank you, your suggestion improves clarity and we have edited accordingly.

6. Results; Page 9; Paragraph 1

“The relative risk of disagreement between chart re-abstraction and administrative data coding when the final diagnosis was uncertain or absent was 1.82 [CI95 1.36 , 2.44]. The risk difference was 18.5%.”

Comment: This information doesn't seem to be located in any of the tables. To clarify, was this a combined RR for “uncertain” or “absent”? If so, please edit the statement to something like..... “As a combined measure, the relative risk of disagreement....” This seems necessary since the tables/figures report them as separate measures (uncertainty vs. lack of definitive diagnosis), yet the text here provides them as a combined measure.

Authors response: You are correct, the relative risk calculation combined uncertain or absent final diagnosis. We have changed the sentence to reflect this (page 9, second paragraph under results):

“As a combined measure, when the final diagnosis was either uncertain or absent, the relative risk of disagreement between chart adjudication and administrative data coding was 1.82 [CI95 1.36 , 2.44], in comparison to those with a clearly documented final diagnosis.”

7. Discussion; Page 10; Paragraph 2

“Physician documentation was the dominant factor influencing the accuracy of the administrative data coding for TIA and minor stroke patients discharged from the ED. The risk difference was 18.5%, suggesting that close to one in five charts with coding misclassifications could have improved accuracy if a definitive diagnosis was documented.”

Comment: Since this is a combined measure (risk difference 18.5%), please change this statement to something like “.... suggesting that coding misclassifications could have been prevented in approximately one in every five charts if a definitive diagnosis was documented or uncertainty removed.”

Authors response: Thank you, we have edited.

8. Results/Discussion

Comment: There is no analysis/discussion regarding comparing accuracy as a whole for the “main position” vs. the “any position” algorithms. This is not your intended purpose of this study, so I don't expect you to perform additional analyses. However, it is very interesting that your re-abstracted TIA cases have higher coding accuracy in “any position” (63.8%) vs. the “main position” (36.8%). I have several theories as to why this might have occurred, but I'd like you to briefly comment on this in the discussion section. I think this is beneficial to discuss since a vast amount of research indicates that codes in the primary position tend to have higher positive predictive values compared to codes in a non-primary position. Yet your results refute this in this instance.

Authors response: Thank you for this comment and the opportunity to discuss. We agree that the diagnostic coding accuracy (sensitivity, specificity, PPV, and NPV) based on code position deserves proper analysis and discussion. We have done these analyses and tested four case definition algorithms against two reference standards. These results have been submitted as a separate manuscript to allow for a more comprehensive discussion.

Reviewer: 3 Reviewer Name: Eric Adelman, MD

Institution and Country: University of Michigan, Ann Arbor, MI USA

Competing Interests: None declared.

The authors aimed to evaluate the accuracy of TIA coding in the ED and examined predictors of accurate coding. The patient data used in these analyses came from 2 TIA trials done at a tertiary care hospital. Charts were abstracted by a vascular neurologist and a second independent review was performed on a subset of charts. TIA was defined by duration of symptoms rather than absence

of brain infarction. The authors found that 39% of patients the stroke neurologists thought had TIAs were not coded at TIAs. Diagnostic uncertainty predicted disagreement between the stroke neurologist review and coder. Ongoing symptoms while the patient was in the ED predicted concordance between the vascular neurology review and coding.

This is an interesting study as administrative data is frequently used in epidemiologic studies and for quality reporting. The methods are appropriate but would have been strengthened by having a coder review the charts in conjunction with the stroke neurologists. Additionally, selection bias is a major limitation of this work. By only including patients enrolled in a TIA trial, the authors enriched their sample with patients at higher likelihood of having a TIA. As the authors note, this lessens the generalizability of the study. There is also a tension between diagnostic uncertainty at the time of presentation and how that uncertainty is documented. The authors seem to presume that a definite diagnosis of TIA can always be made in the ED. Perhaps it is easier to make a TIA diagnosis in retrospect and this would impact the authors' conclusions. The recommendations about improving documentation when there is diagnostic uncertainty are well said.

Specific comments:

Strengths and limitations section

--Line 6: Consider reframing the first bullet point. If the administrative data is not good at identifying TIAs in the ED, data accuracy is not particularly relevant. Perhaps putting the emphasis on predictors of data accuracy as a way to improve the sensitivity of administrative data would emphasize this point.

Authors response: Thank you for your review. We modified the first bullet to emphasize the relevance in improving data accuracy (page 3):

“Understanding the predictors of data accuracy has high relevance in improving the low to moderate sensitivity of transient ischemic attack coding in the emergency department and may shed insight into the coding accuracy of other conditions primarily treated in the outpatient setting.”

--Line 20: The authors refer to the “patient-physician” interaction, but it is not clear what this means. Authors response: “Patient-physician interaction” refers to elements that may influence patient care, but are not specifically a patient or physician characteristics, e.g. decision for advanced neuroimaging, specialist involvement, and time of presentation to hospital. These decisions are often the result of the interaction that occurred or its setting. To clarify, the third bullet was edited (page 3):

“We evaluated predictors related to patients, physicians, and the patient-physician interaction, such as decision for advanced neuroimaging, consultation of a specialist, and time of presentation, including factors that can be acted upon to improve administrative data accuracy.”

--Consider noting that this study was done in patients participating in a TIA trial. That seems to be a bigger limitation than the fact it was done at a university teaching hospital.

Authors response: We agree this is an important limitation to state. To clarify, the fourth bullet was edited (page 3):

“The study population was obtained from two observational studies on the diagnosis of TIA and minor stroke versus stroke mimic using serum biomarkers and neuroimaging and most patients were discharged from a single university teaching hospital, limiting the generalizability of our results to other jurisdictions and healthcare models.”

Introduction

--Page 4, line 10: These are patients with TIAs discharged from the ED, not TIAs discharged from the

ED.

Authors response: We have edited to clarify this sentence.

--Page 4, line 19: It is not clear what the authors mean by "clinical acumen"

Authors response: We have changed the word "acumen" to "judgment" to illustrate the subjective nature of making the clinical diagnosis of TIA.

--Page 4, line 29: If the authors are going to highlight physician factors, be specific about what factors those are. Reading this, one thinks of role (trainee, APP, attending) or time in practice, as physician factors but the authors do not have this information.

Authors response: Thank you for this comment. The physician factors studied include specialist versus emergentologist evaluation and content of documentation. We agree that other physician characteristics, such as role, training background (e.g. IMG vs not, general neurologist vs stroke specialist, ED family physician vs emergentologist), and time in practice are of relevance. Due to privacy concerns, we were unable to obtain this information. Further, our limited sample size is not powered to fully study the large number of relevant factors that need consideration. We revised the sentence to read (page 4, last paragraph):

"We aimed to study the associations between patient characteristics and physician factors, including physician specialty and chart documentation,..."

Methods

--Page 5, line 25: The authors should highlight why they used a time based definition of TIA, rather than the contemporary tissue based definition (Stroke 2009;40:2276-2293)?

Authors response: We agree this is an important point to discuss and have given the two definitions much consideration. We have added (page 13, second paragraph):

"Using a tissue-based diagnosis for TIA²⁹ based on an objective biologic endpoint has potential to improve the inter-rater reliability. However, cerebral ischemia can only be demonstrated in 30-70% of clinically-diagnosed TIA using magnetic resonance diffusion-weighted imaging and the finding of ischemia is dependent on symptom duration.³⁰ Further, because diffusion-weighted imaging is not routinely utilized in clinical practice, applying a tissue-based TIA diagnosis as reference standard would negatively impact the generalizability of our results."

--In the methods section the authors may want to comment that they were unable to adjust for ED and neurology provider characteristics, specifically role (attending, APP, resident, stroke specialists, etc) and experience (e.g. years in practice.)

Authors response: This is indeed a limitation of our methods and we added in the limitations paragraph (page 13, last paragraph):

"We had limited information on individual physician or coder characteristics, such as their role (trainee, attending, subspecialty), level of training, and experience."

Discussion

--Page 10, line 31: It seems to be valid that ongoing symptoms and neurology consultation improve coding, but it is not clear how male sex impacts coding accuracy.

Authors response: Thank you for the opportunity to discuss. We can propose different theories to explain the association between male sex and improved coding accuracy. It is possible that physicians interpret male patients' symptoms more seriously than symptoms from female patients,

that male patients are more likely to present to ED with more serious or prolonged symptoms; or that male patients describe their symptoms in a more textbook fashion, e.g. “my arm was weak and I couldn’t move it” versus vague symptoms, e.g. “my body felt tired and I wasn’t coordinated.” The true explanation is likely multifactorial. We do not have data to support these theories and do not want to speculate in our discussions.

--How much knowledge do physicians have of coding? The authors talk obliquely about this, but their work would be strengthened by addressing this issue more directly. If physicians aren’t knowledgeable about coding requirements then it is not surprising that these requirements are not met. In a similar vein, more concrete recommendations on how to improve coding would be useful. Authors response: Thank you for this comment. We have added the following text and references (page 12):

“Specifically, data from the US demonstrate that physicians do not receive adequate training in disease nosology and coding methods.²¹ In certain European countries, physicians are responsible for administrative data coding, but the degree of formal training is variable. An Italian study evaluating physicians’ inter-coder agreement for stroke codes defined an “expert coder” as a neurologist who underwent at least three half- or full-day ICD-coding training courses.²² In comparison, Canadian coding specialist are trained in a two-year post-secondary education program.²³”

Conclusions

--The manuscript would be strengthened if the authors were more specific about their findings in this section.

Authors response: As suggested in this comment, the conclusions have been updated (page 14):

“Clear documentation of a final diagnosis was associated with improved administrative data accuracy for TIA and minor strokes in the ED. Although our data do not reveal the reason why one chart is better coded than another, an intuitive solution is to educate physicians and other providers who generate medical information about proper documentation techniques, disease nosology, and coding standards. Prospective studies evaluating the effects of techniques to improve physician documentation, such as education programs, standardized discharge summaries, or automated coding algorithms, on coding accuracy are needed. Administrative data are continuously generated – at a cost to the healthcare system – to help decision-makers monitor, evaluate, and plan the provision of health services as well as for research. Improving their accuracy is in the best interest of patient care.”

Tables

Table 2: consider adding another column with the results of an appropriate statistical test of difference between agreement and disagreement.

Authors response: The test of proportion comparison of the predictor variables of interest in Table 2 is equivalent to the results of the univariable analysis in Table 3. Given we were interested in two TIA algorithms (main-position vs any-position), we did not want to highlight the comparison of proportions in Table 2.

Reviewer: 4 Reviewer Name: Cheng-Yang Hsieh

Institution and Country: Department of Neurology, Tainan Sin Lau Hospital, TAIWAN

Competing Interests: None declared

This study aims to explore the relationship between physician documentation and the accuracy of administrative data coding in TIA patients. In general, the manuscript is well-written. And I totally agreed that physician documentation is very important for coding accuracy in paroxysmal disorders

such as TIA. I have the following comments for the authors in the future revision:

1. Title: Is this really a COHORT study? Why?

Authors response: Thank you for your review and comments. We determined our study to be a cohort study because it is an observational study of outcomes from individual patients, as opposed to aggregates. We analyzed the association between different exposures (age, sex, ABCD2 score, consultation by a neurologist, content of physician documentation, etc) and the outcomes (agreement vs disagreement of the diagnosis between coders and chart adjudicators). By definition, the exposures occurred before the codes were determined. The study has a forward direction of logic, but data were retrospectively collected.

2. Line 54 of Page 4: “Both studies require at least one clinical evaluation by a neurologist”. But in Table 2, 75.4%-84.4% of patients had neurology consult, not 100%.

Authors response: You are correct, our sentence was unclear. Both studies required a neurologist evaluation either in the ED or at least once in the clinic as follow-up. We have revised to remove this sentence because the information is not directly relevant for our objectives.

3. Line 26 of Page 5: The authors adopt the WHO time-based criteria for diagnosing TIA when chart re-abstracted. For a patient with symptoms lasting less than 24 hours but a MRI showing evidence of tissue ischemia, the diagnosis would be adjudicated as a TIA, but not a stroke. Will the different definition of TIA (WHO time-based vs. MRI-based) affect results of the present study? Will the administrative coders give the ICD-10 code of cerebral infarct just based on the MRI results?

Authors response: We would like to address comments #3 and #4 together. We have added on page 13, second paragraph:

“In the subset of charts that underwent a second independent review, we showed that the agreement was 76.7%. This is consistent with a study conducted in the Stanford TIA clinic that reported a 72% overall agreement among fellowship-trained stroke neurologists.²⁸ The subjectivity in TIA diagnosis, especially during the initial health encounter, is impossible to eliminate. Using a tissue-based diagnosis for TIA²⁹ based on an objective biologic endpoint has potential to improve the inter-rater reliability. However, cerebral ischemia can only be demonstrated in 30-70% of clinically-diagnosed TIA using magnetic resonance diffusion-weighted imaging and the finding of ischemia is dependent on symptom duration.³⁰ Further, because diffusion-weighted imaging is not routinely utilized in clinical practice, applying a tissue-based TIA diagnosis as reference standard would negatively impact the generalizability of our results.”

4. Line 31 of Page 5: A senior stroke neurologist independently reviewed a random sample of thirty charts to determine inter-abstractor reliability. And in Line 45 of Page 5, the agreement was only modest (76.7%, kappa = 0.50). How to explain this disagreement between two stroke neurologists? Given such a modest agreement of TIA diagnosis between stroke neurologists, it may not be surprising for low agreement between physician document and administrative coders in diagnosis of TIA.

Authors response: Please see response above.

5. Line 45 of Page 6: “a 85-year-old woman...” It should be a typo (“82F” in Table 1).

Authors response: Thank you, we edited the text.

6. Discussion: Will other factors, such as insurance’s policy for reimbursement, affect the coding behavior? For example, if the reimbursement ED visits for ischemic stroke is much higher than ED visits for TIA, will the coder more likely to code ischemic stroke?

Authors response: Thank you, this is a good point. We added the following to the limitations (page 14, paragraph before Conclusions)

“Finally, although patient diagnosis does not influence physician reimbursements in Canadian acute care hospitals, payment systems and financial incentives can influence diagnosis coding³². We did not have access to longitudinal payment information to test this hypothesis.”

VERSION 2 – REVIEW

REVIEWER	Jonathan Thigpen Notre Dame of Maryland University School of Pharmacy, USA
REVIEW RETURNED	27-Feb-2017

GENERAL COMMENTS	<p>Thank you for your revisions. The authors addressed most of my comments. However, I’m still concerned with the clarity of Figure 1 (and its related text in the paper). I spent a lot of time trying to decipher the results related to Figure 1. I approve this paper pending revision of Figure 1 and its related text.</p> <p style="text-align: center;">Review</p> <p>A cohort study on physician documentation and the accuracy of administrative data coding to improve passive surveillance of transient ischemic attacks</p> <p>Yu et al.</p> <p>Thank you for your revisions. The authors addressed most of my comments. However, I’m still concerned with the clarity of Figure 1 (and its related text in the paper). I spent a lot of time trying to decipher the results related to Figure 1. I approve this paper pending revision of Figure 1 and its related text.</p> <p>Please address my comment below.</p> <p>1. Results; Page 9</p> <p>“Among 417 patients included, the re-abstracted adjudicated diagnoses showed 163 (39.1%) TIA, 155 (37.2%) mild ischemic strokes, and 99 (23.7%) stroke mimics. Inter-abstractor agreement between the two stroke neurologists was 76.7% ($\kappa = 0.50$) for the diagnosis of TIA. The main-position NACRS algorithm identified 79 cases of TIA and the any-position algorithm identified 151 cases. Figure 1 shows the distribution of cases of disagreement between chart re-abstraction adjudication and the two TIA coding algorithms. Among the 79 patients coded with TIA in the main position, there were 122 cases of disagreement for the main-position algorithm (shaded in grey, n=60 confirmed and n=103 denied). Among the 151 patients coded with TIA in any position, there were 106 cases of disagreement (unshaded, n=104 confirmed and n=59 denied). Finally, the cerebral ischemia algorithm identified 202 cases of TIA and minor stroke and there were 138 cases of disagreement for this diagnosis (Figure 2).”</p> <p>Comments: There is a total of 163 confirmed/adjudicated TIA cases. Of those, 60 were coded in the main position and 103 were not coded as TIA at all. The abstract states “The majority</p>
-------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

of disagreement (n=103, 84.4%) arose from adjudicated TIA cases that were misclassified as non-TIA in administrative data coding. Is this correct? If so, shouldn't these 103 cases also be within the "NACRS any-position incorrectly coded" cell as well (n=59)? It seems that if a confirmed TIA case was misclassified and coded as non-TIA, then it would be included in both the "incorrectly coded" main position and any position cells. Essentially, how are the "incorrectly coded" cells different for the "main position" and "any position"? Please clarify and then revise text for the abstract and results as needed.

Also, there are other clarifications needed for cells in Figure 1. For instance, what does "incorrectly coded" refer to? Does it mean a TIA case was not coded as TIA when in fact it should have been? Does it mean that a patient was coded for TIA when in fact they should NOT have been? Or does it mean that the patient was correctly coded for TIA, however the code was in the wrong position?

Another example of the issues with Figure 1 is the following statement from the text "**Among the 79 patients coded with TIA in the main position, there were 122 cases of disagreement for the main-position algorithm (shaded in grey, n=60 confirmed and n=103 denied).**" How can it be that among 79 patients, there are 122 cases? (numerator bigger than denominator?) Also, n=60 + n=103 equals n=163 which refers to the number of total adjudicated confirmed TIA cases. However, here it says the n=103 is "denied"? What does that mean? Does that mean they weren't adjudicated and confirmed as TIA? Also, shouldn't the text instead be "**122 cases of disagreement for the main-position algorithm (shaded in grey, n=103 confirmed and n=19 denied).**"

In general, I recommend the authors revise the results section to describe (in detail) the various components of Figure 1. Also, make sure that the numbers add up. If describing the results and the numbers listed don't add up, it is very confusing for readers and difficult to follow along. The authors may want to consult other ICD accuracy papers to get a better sense of how to clearly report their findings. Below is a reference they may want to use.

Thigpen JL, Dillon C, Forster KB, et al. Validity of International Classification of Disease Codes to Identify Ischemic Stroke and Intracranial Hemorrhage among Individuals with Associated Diagnosis of Atrial Fibrillation: Thigpen et al: ICD-9 Stroke Case and Atrial Fibrillation Accuracy. *Circulation Cardiovascular quality and outcomes*. 2015;8(1):8-14. doi:10.1161/CIRCOUTCOMES.113.000371.

REVIEWER	Eric Adelman University of Michigan Ann Arbor, MI USA
REVIEW RETURNED	16-Feb-2017

GENERAL COMMENTS	<p>I appreciate the authors response to my comments. I have a few additional minor comments.</p> <p>On page 13, lines 6-20: Using the tissue based definition of stroke, when there is DWI change or evidence of infarction on CT in a brain region that corresponds to clinical symptoms the patient had a stroke. A TIA is when there is no evidence of infarct or tissue damage. It is not clear what the authors are trying to say as a tissue based diagnosis of TIA means transient symptoms attributable to cerebral ischemia without evidence of infarction. This section would benefit from clarification.</p>
-------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

REVIEWER	Cheng-Yang Hsieh Department of Neurology, Tainan Sin Lau Hospital, TAIWAN
REVIEW RETURNED	17-Feb-2017

GENERAL COMMENTS	The authors clearly and appropriately addressed my comments. Thanks for their effort in this important study. I also learned a lot during this review process.
-------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------

VERSION 2 – AUTHOR RESPONSE

Revision 2: Response from Authors to Reviewers' Comments:

Reviewer: 2

Reviewer Name: Jonathan Thigpen

Institution and Country: Notre Dame of Maryland University School of Pharmacy, USA

Competing Interests: none declared

See file attached.

Thank you for your revisions. The authors addressed most of my comments. However, I'm still concerned with the clarity of Figure 1 (and its related text in the paper). I spent a lot of time trying to decipher the results related to Figure 1. I approve this paper pending revision of Figure 1 and its related text.

Thank you, we appreciate the importance of clarifying Figure 1. We have separated it into two figures (Figure 1 and 2) and modified the text accordingly:

Figure 1 shows the distribution of cases of disagreement between chart adjudication and the NACRS main-position algorithm for the diagnosis of TIA. Among 163 chart-adjudicated cases of TIA, there were 60 (36.8%) charts that were correctly coded as a TIA in the main-position and there were 103 (63.2%) charts that were incorrectly coded as non-TIA. Among 254 chart-adjudicated cases of non-TIA (minor strokes or mimics), there were 235 (92.5%) charts that were correctly coded as a non-TIA in the main-position and 19 (7.5%) charts that were incorrectly coded as a TIA. The total number of incorrectly coded charts was 122. Most of the incorrect coding was explained by cases of chart-adjudicated TIA that were incorrectly coded as non-TIA (103/122 84.4%). Figure 2 shows the cases of disagreement between chart adjudication and the NACRS any-position algorithm.

Reviewer: 3

Reviewer Name: Eric Adelman

Institution and Country: University of Michigan, Ann Arbor, MI, USA

Competing Interests: None declared

I appreciate the authors response to my comments. I have a few additional minor comments.

On page 13, lines 6-20: Using the tissue based definition of stroke, when there is DWI change or evidence of infarction on CT in a brain region that corresponds to clinical symptoms the patient had a stroke. A TIA is when there is no evidence of infarct or tissue damage. It is not clear what the authors are trying to say as a tissue based diagnosis of TIA means transient symptoms attributable to cerebral ischemia without evidence of infarction. This section would benefit from clarification.

Thank you, we have clarified the paragraph:

The subjectivity in TIA diagnosis, especially during the initial health encounter, is impossible to eliminate. Demonstration of cerebral ischemia using magnetic resonance diffusion-weighted imaging in clinically-diagnosed TIA patients is variable (30 to 70%) and associated with longer symptom duration. A tissue-based diagnosis for TIA has been proposed, where the diagnosis is defined by the absence of tissue ischemia. However, because diffusion-weighted imaging is not routinely utilized in clinical practice, applying a tissue-based TIA diagnosis as reference standard would negatively impact the generalizability of our results.

Reviewer: 4

Reviewer Name: Cheng-Yang Hsieh

Institution and Country: Department of Neurology, Tainan Sin Lau Hospital, TAIWAN

Competing Interests: None declared

The authors clearly and appropriately addressed my comments. Thanks for their effort in this important study. I also learned a lot during this review process.

Thank you.

VERSION 3 – REVIEW

REVIEWER	Jonathan Thigpen Notre Dame of Maryland University School of Pharmacy United States
REVIEW RETURNED	06-Apr-2017

GENERAL COMMENTS	Thank you for addressing my second round of comments. You've clarified the results/discussion regarding Figure 1 and case validation. Thank you.
-------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------