Reviewer's report

Title: The University of Texas Houston Stroke Registry (UTHSR):
Implementation of Enhanced Data Quality Assurance Procedures Improves Data
Quality

Version: 1 Date: 4 November 2012

Reviewer: Mathew J Reeves

Reviewer's report:

Given the limited data published regarding the accuracy of data collected by
stroke registries this manuscript – which describes the development and
assessment of data assurance mechanisms for the University of Texas Stroke
Registry - is potentially a useful contribution.

The main limitations of the manuscript are: 1) its excessive length, 2) limited
sample size of the reliability and validity assessments, 3) standard validity
measures of Sensitivity and Specificity were not utilized. More detailed
comments follow. Major Compulsory Revisions [MCR] are indicated as such in
bold. All other suggestions are discretionary.

Introduction:

If this section is designed to be a comprehensive review of the history of stroke
registries then some editing is required:

- The Riks Stroke registry from Sweden deserves mention. [MCR]
- The PCNASR now involves more than the 4 states listed and the original pilot
  registry included OH and MI (See Stroke 2005; 36:1232-1240). [MCR]
- Other validation/reliability studies (apart from ref 12 and 26) relevant to stroke
  registries include:
  o Xian et al. Am Heart J. 2012 Mar;163(3):392-8.,
  o Reeves et al., Prev Chronic Dis. 2011 May;8(3):A62.

Methods:

Page 14: What is the number of individual data fields for the registry? How was
this database developed – was it based of prior registries like PCNASR?, or
GWTG? or was it developed de novo?

What about post-discharge outcomes information? – this was briefly mentioned
but no details were given.

The ability for EMR to capture data elements is obviously an important
enhancement – what % of the data fields are captured electronically and what
medical information does it cover?

Page 17: The sample size for both the reliability study (n= 30) and validity study
(n= 41) are quite small. Or is it that these data are repeated every year? Why were such numbers chosen and how do such small numbers impact power calculations for both kappa and ICC? [MCR]

Page 18: The section on the Assessment of data quality could also be referred to as a validity study since the primary analysis was to compare the data to a gold standard. More details could be provided on the clinician adjudication process (was this one clinician?) Clarify here that the 41 records used in this process are different from the 30 and 13 records mentioned on page 17 under the data training section? The error rate was calculated amongst all entries re-abstracted – how many was this? Also, why wasn’t Sensitivity and Specificity data calculated since there is a gold standard? Kappa and ICC should really be only used for reliability studies. [MCR]

No mention in the methods of the use of the stroke quality metrics shown in Table 4.

Results:

Page 21: So the annual error rates (08-11) are based on only about 10 records each? It seems to me that confidence intervals should be added to these figures (including those shown in Figure 2). [MCR] Also is this really a “trend” or do the data simply show improvement between 08 and 09 and then a new low error rate for the 09-11 period?

Page 22: The air transport figure is very high – is this because of the equally high transfer rate? Is there any evidence of a statistical “trend” in the data shown in Table 3? Also, why not show the annual trends in the data in Table 4? Also please add footnotes to the table to describe the definitions used for each measure. [MCR] This would be the first study to ever report 100% compliance with the STK-3 measure……

Page 23: Is tPA time onset to needle time or door to needle time? [MCR] The denominators for the measures calculated in Table 5 should be shown. [MCR] Also interpretation of the negative or positive bias index when the gold standard is being used requires knowledge of who is being compared to whom (what does a positive bias mean?). Explain why the measures with perfect agreement (Kappa = 1.0) have no variability in their 95% CI’s – does it make sense to calculate the CI in this situation?

Discussion

Please provide some information on why you think the reliability/ accuracy data were not perfect for measures such as hospital arrival and tPA time. What is going on in the registration process at the hospital that might be contributing to these problems? Is tPA time affected by the known difficulty of assessing stroke onset time? [MCR]

**Level of interest:** An article whose findings are important to those with closely related research interests
**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**
I declare that I have no competing interests.