Reviewer's report

Title: Hospital Volume and Mortality in Non-elderly Pneumonia Patients Stratified by Severity: a Retrospective Cohort Study

Version: 1 Date: 27 September 2013

Reviewer: Henry J Carretta

Reviewer's report:

1. Major Compulsory Revisions-None.
2. Minor Essential Revisions
   a. Improve the completeness of the table and figure labels so that the content is understandable without reading the text. For example, Table 1. “Patient characteristics overall and by hospital volume for discharges with a primary diagnosis of pneumonia, Japanese Diagnostic Procedure Combination database, 2010” or Table 2. “Unadjusted and adjusted logistic regression models for the likelihood of inpatient mortality associated with community acquire pneumonia, Japanese Diagnostic Procedure Combination database, 2010”. All tables and figures need improved titles.
   b. Create an additional descriptive table that provides more information on the variability of volume, deaths, and mortality rates across all hospitals. I am not asking for individual hospital rates to be published but summary measures of the variability in volume deaths and mortality rate (which should be moved from Table 1 to the new table) across hospitals.
   c. Why were those cut points chosen for the hospital volume categories?
   d. Are there limitations due to the 6 month period of data capture? Provide supporting citations for whether community acquired pneumonia in this age group varies by season or not and how that might influence findings.
   e. Are there limitations to inference associated with the small number of discharges per hospital and small number of deaths? Should future research pool multiple years to improve estimation? What if any are the analytic issues when using GEE when so many hospitals have a small number of discharges and zero or only a few deaths. For example, are ten discharges adequate to create an appropriate correlation structure?
   f. What are the limitations associated with only capturing 45% of discharges nationally? What is known about hospitals that do not report to DPC? To what degree can it be said that the reporting hospitals are representative or not representative of all hospitals?
3. Discretionary Revisions
   a. Logistic regression is the usual approach but I wonder if a count model like Poisson or negative binomial might provide some additional insight. Volume is the principal interest but a count model with volume as an indicator versus a
count model with volume as the offset or exposure might provide additional information and would allow use of a zero inflation approach to account for so many hospitals with no deaths. The later would be a good way to characterize differences in the population with volume held constant as a starting point. This would also be a convenient way to descriptively characterize unadjusted variation across hospitals as noted in 2 b.

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

'I declare that I have no competing interests’