Reviewer's report

Title: Further Validation that Claims Data are a Useful Tool for Etiologic Research on Hypertension

Version: 1 Date: 12 September 2012

Reviewer: Guanmin Chen

Reviewer's report:

General comments:
The author sought to investigate hypertension validation for claims data using blood pressure reading. Based on a large cohort study population, this paper compared the hypertension status between health screening data and administrative data, and provided an important epidemiologic issue for using administrative data.

Major compulsory revisions

1. Definition for hypertension in health screening data: the hypertension status (standard) was only defined using blood pressure reading. For diagnosed hypertension with antihypertensive medication, I concerned the blood pressure readings are normal due to controlled hypertension. This means the diagnosed hypertension, but controlled with medication are missing. Except for the medication, the history of hypertension, physician’s notes etc. should be included for the hypertension definition. Although the author discussed this issue, such as the medications are used for other condition, not for hypertension, providing the sensitivity analysis results will be a support for this.

2. Definition for hypertension in administrative data: physician billings and/or hospitalization using ICD-9 coding. Please clarify how many diagnosis codes for physician billings, for hospitalization admissions. The hypertension was identified using first(primary) diagnosis code or all possible diagnosis. For the definition “at least two separate medical claims” do you have any time gap requirement for the two claims for example, two claims time gap is less than 1 years, or 2 years, etc.

3. Please clarify how many times(round) for the health screening were performed during study period, and how many round heath screening data were used for the analyzing.

Minor Essential Revisions

1. For Monte Carlo logistic regression, the hypertension was identified using claims; please clarify which definition you use for hypertension: 1) at least one medical claims; or 2) at least two separate medical claims.

2. For restrictive cases definition, please clarify how to define “consistently normal or hypertensive”, what’s the operational definition for “consistently”, such
as time periods and time gap for 2nd and 3rd reading, etc.

Discretionary Revisions

1. Emphasize the issue for the hypertension missing claims as comorbidities, such as the diabetes with hypertension, and the physicians billings only claims diabetes,
2. Narrow down the methods section “study company”
3. I am not clear the cohesion of “etiological research on hypertension” to this paper, please clarify.

Level of interest: An article of importance in its field

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.