Statistical Method Supplement

Description: Supplemental statistical method material for Pneumonia Hospitalization Coding Changes associated with Transition from the 9th to 10th Revision of International Classification of Diseases

1. Introduction

When a disease is diagnosed by different diagnostic algorithms in different populations, or in the same population at different time points, the reported disease incidence rates cannot be compared directly because of differences in accuracy of the diagnostic algorithms. If we take one population, or time point, as the standard or baseline, then incidence rates of the other populations, or time points, need to be adjusted according to the relative accuracy of the corresponding algorithms at baseline. To account for changes in diagnostic coding systems used to identify patients hospitalized with pneumonia, we developed a simple, yet general method, for such adjustment. We describe the method as well as its implementation details in this short statistical supplement.

2. Main results

Because of the transition from ICD-9-CM to ICD-10-CM in October 2015, we have two periods P_b (pre- ICD10-CM implementation) and P_a (post- ICD10-CM implementation) in our study using two different diagnostic algorithms A_b (ICD-9-CM algorithm) and A_a (ICD-10-CM algorithm) used to identify the same disease *D* (pneumonia). Although the approach is designed to allow comparisons of rates over time and encompassing the transition of ICD coding systems, for the derivation of the

estimates we assumed that the underlying population remained generally constant during the two periods and that there were no changes in the true incidence of the disease during the transition of coding systems. Our objective was to derive an adjustment factor that could be applied to correct for the impact of the coding system transition. Thus, using r_b and r_a as the corresponding observed incidence rates for each period, the application of our adjustment factor to r_a would make it comparable to r_b.

A diagnostic algorithm's accuracy is summarized by its sensitivity and specificity.[1-3] Considering algorithm A_b as the gold standard, the relative accuracy of A_a relative to A_b can be defined by following two quantities:

$$\theta = P(A_a = 1 \,|A_b = 1) \qquad [1]$$

$$\delta = P(A_a = 1 \,|A_b = 0) \qquad [2]$$

Here each algorithm takes the value of 1 or 0 if the disease was either diagnosed by the algorithm or not. θ and $1 - \delta$ are the sensitivity and specificity of A_a relative to A_b, respectively.

If we know both θ and δ then the disease rate r_a from applying algorithm A_a can be adjusted to $\hat{r}_a$, which is comparable to what would be produced if the gold standard algorithm A_b was applied. $\hat{r}_a$ can be easily calculated by the following formula:

$$\hat{r}_a = \frac{r_a - \delta}{\theta - \delta} \qquad [3]$$

For example, if A_a is identical to A_b, then $\theta = 1$ and $\delta = 0$, hence $\hat{r}_a = r_a$, and no adjustment is needed.

To estimate the sensitivity, one can take a random sample of patients diagnosed by the gold standard algorithm A_b (i.e., $A_b = 1$), and then apply algorithm A_a to the sample. The frequency of pneumonia captured by the A_a algorithm would be an unbiased estimate of the sensitivity θ.

While, in principle, we can estimate the specificity $1 - \delta$ in a similar manner by taking a random sample of subjects that do not have pneumonia from applying algorithm A_b (i.e., $A_b = 0$), and then apply algorithm A_a to the sample, and count the number of patients without pneumonia, it is not practical for most diseases because of the usually low prevalence rate of the disease in study population. We would have to apply both algorithms to a large number of healthy subjects in order to obtain a reasonable estimation of the specificity.

On the other hand, it is relatively simple to estimate the positive predictive value (PPV) of A_a relative to A_b. Here PPV is defined as:

$$PPV = P(A_b = 1 \,|A_a = 1) \qquad [4]$$

Namely, the proportion of pneumonia captured by the gold standard algorithm A_b among those who were captured as pneumonia by algorithm A_a [2-3]. Unlike the specificity, $1 - \delta$, PPV can be estimated easily by taking a sample of patients that are diagnosed as positive by A_a, then applying A_b to the same sample and count the number classified as pneumonia. In fact, our study was designed such that PPV and the sensitivity θ can be estimated from two samples of 500 pneumonia cases each

captured by the ICD-9-CM algorithm (A_b) and the ICD-10-CM algorithm (A_a), respectively.

If we also know the incidence rate r_b from applying the gold standard algorithm A_b, then there is an established relationship between PPV and specificity[2]:

$$PPV = \frac{\theta\, r_b}{\theta\, r_b + \delta\,(1 - r_b)}$$

We can then calculate the specificity $1 - \delta$, with

$$\delta = \theta\, \frac{r_b\,(1 - PPV)}{(1 - r_b)PPV} \qquad [5]$$

Should we apply both algorithms in the same population, then the adjusted rate $\hat{r}_a$ derived from algorithm A_a should equal to the unadjusted rate r_b derived from algorithm A_b. Therefore, from equation [3] and [5] together, we obtain a simple, yet general formula to calculate the adjusted incidence rate as in the following Lemma:

Lemma: If θ, the sensitivity of A_a relative to A_b and PPV, the positive predictive value, as defined in Equation [1] and [4], are both known, then the incidence rate r_a from applying A_a can be adjusted by applying the following formula:

$$\hat{r}_a = r_a\, \frac{PPV}{\theta} \qquad [6]$$

To be comparable with incidence rates from the gold standard algorithm A_b, we can use $\frac{PPV}{\theta}$ as the adjustment factor.

3. Implementation of the method:

As discussed in the paper, there is substantial variability in applying the two algorithms, along with the sampling variability of selecting patients. Such variability is reflected by the different sensitivity and PPV values. Some variability is systematic, such as the difference between children and adult patients, while other variability appears random, such as variability among coders. We follow general epidemiology and statistical practices to deal with these variabilities. For systematic variability, we stratified the analysis by the systematic factors. Hence, our results are analyzed and reported by children and adults separately.

We could consider coders as a random effect if we had a relatively large number of coders (say greater or equal to 5). However, in our study, we had three coders for children and four coders for adults. Hence, it was not possible to construct a random effect model to estimate the sampling distribution of the sensitivity and PPV. Instead of that, we considered the means of sensitivity and PPV to be distributed uniformly in the ranges of the observed sensitivity and PPV values. For given means and sample sizes, the actual sensitivity and PPV values were considered as samples from binomial distributions with the means and sample sizes, which in turn could be approximately by normal distributions for large and moderate sample sizes. Then we applied the following procedure to generate 10,000 values of the adjusted factor:

Step 1: Stratified by children and adults, obtain corresponding ranges of sensitivity and PPV.

Step 2: Repeat 10,000 times:

Step 2.1: Sample uniformly from the ranges to get mean values of sensitivity and PPV

Step 2.2. Use the mean value and sample size 500, to calculate standard errors of the observed sensitivity and PPV.

Step 2.3. Sample from normal distributions with above mean and SE to obtain a realization of sensitivity and PPV.

Step 2.4. Calculate the adjustment factor based on the realization.

Step 3: Calculate summary statistic of the adjustment factors based on the 10,000 values.

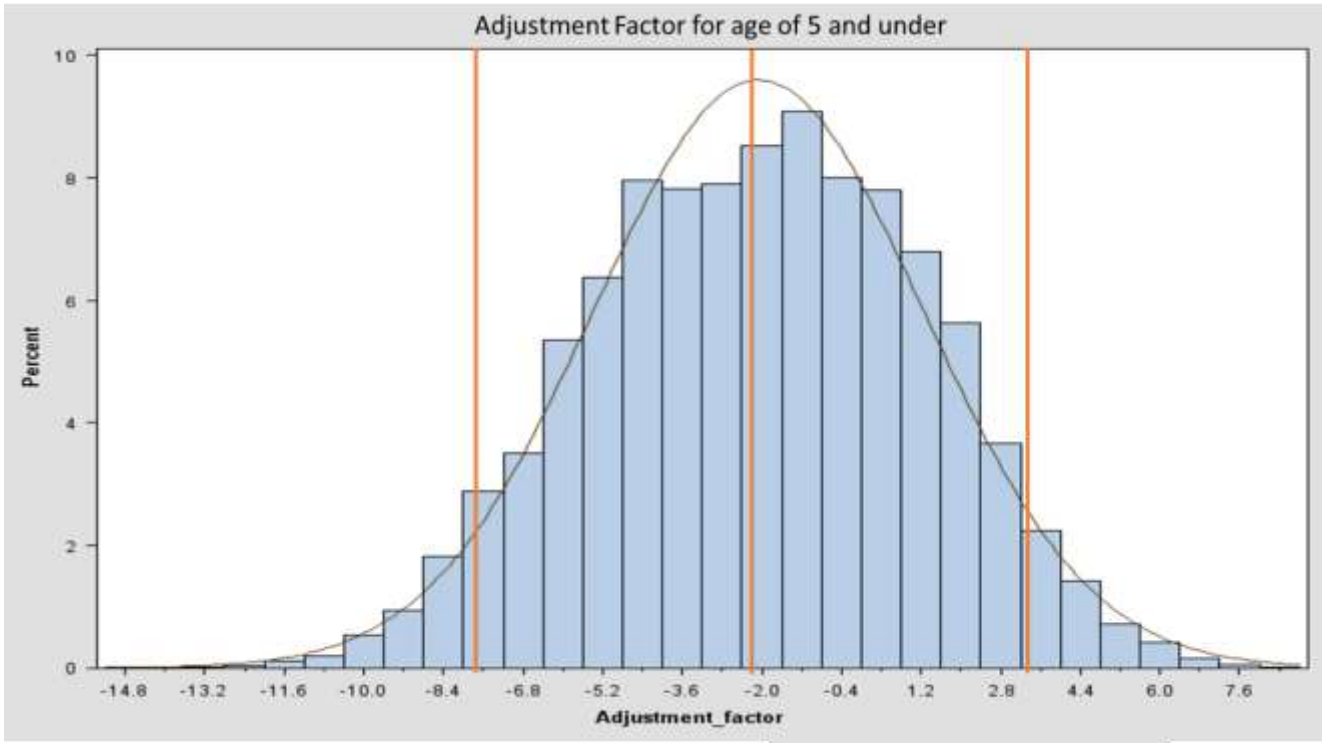
In the main paper, we reported the mean and standard deviation (translated to credible region) of the 10,000 values.

4. References:

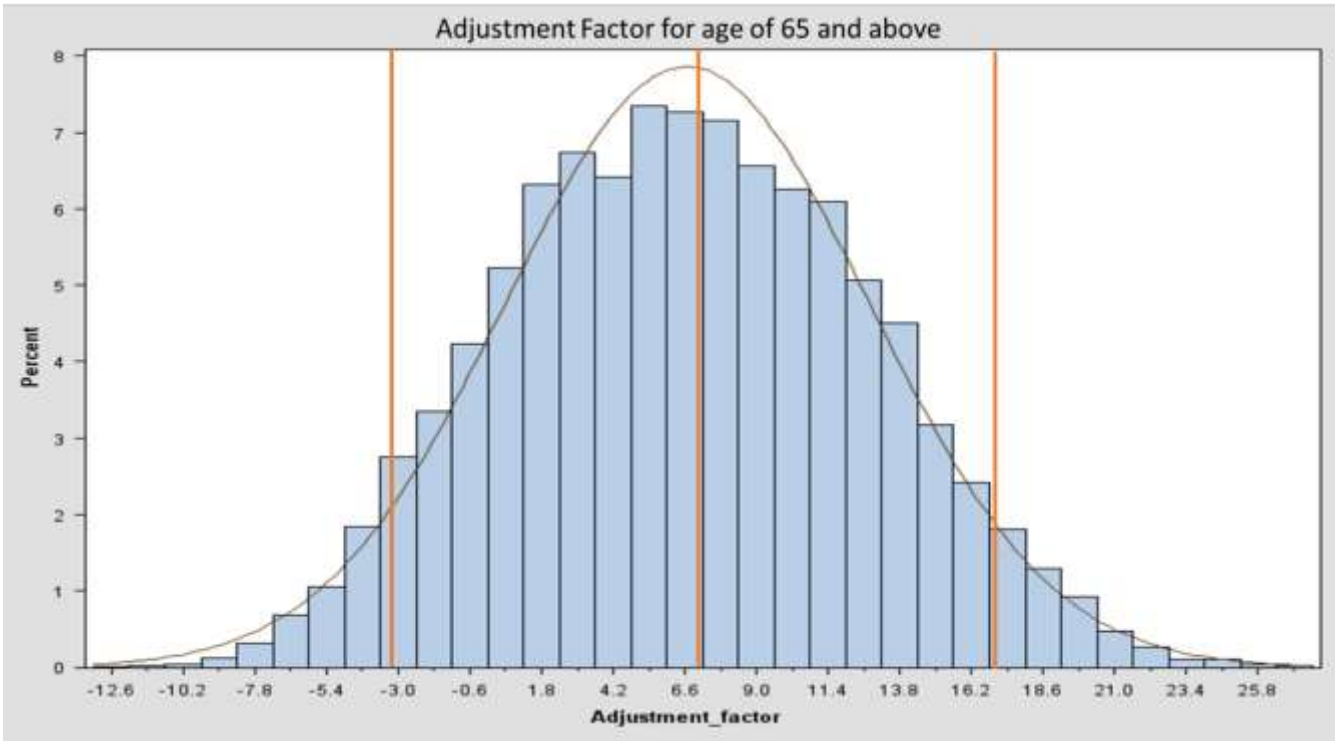
1. K. Chu, An introduction to sensitivity, specificity, predictive values and likelihood ratios. *Emergency Medicine.* 1999; 11: 175-181
2. R. Simon, Sensitivity, specificity, PPV, and NPV for predictive biomarkers. *Journal of National Cancer Institute.* 2015; 107(8): djv153
3. M. Stojanovic, et. al., Understanding sensitivity, specificity and predictive values. *Vojnosanitetski Pregled.* 2014; 71(11): 1062-1065

Figure S1: Adjustment factor for pneumonia hospitalization rates using a discharge code algorithm[2] for children (a) and adults (b) after transition to ICD-10-CM

a) Children ≤5 Years of Age



b) Adults ≥65 years of Age



Vertical lines represent mean and 95% credible region

156

157

158