Author's response to reviews

Title: Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research

Authors:

Alexandar Rusanov (ar2765@columbia.edu)
Nicole G Weiskopf (ngw2105@columbia.edu)
Shuang Wang (sw2206@columbia.edu)
Chunhua Weng (cw2384@columbia.edu)

Version: 4 Date: 28 May 2014

Author's response to reviews: see over
Dear Editor,

Thanks for the opportunity to further improve this work. In this revision, we inserted a summary of negative binomial repression as suggested by the editor and the reviewer two. Below are our detailed responses to the previous review comments.

**REVIEWER 2**

Reviewer: Jose Alberto A Maldonado

**Comments:** “Discretionary revision: I still miss a more detailed explanation of the statistical method, in this new version of the paper a negative binomial regression.”

**Response:** Thank you for this suggestion. We expanded the methods section by adding the following paragraph to summarize this method:

The variations of the counts of laboratory results and medication orders are far greater than the means. To account for this over-dispersion we fit a negative binomial regression.[36] While the Poisson regression (which uses a Poisson distribution) is commonly used for analyses of count data, it does not handle over-dispersed data sets well due to the assumption that the variance of counts equals the mean. The negative binomial regression is an extension of the Poisson regression that is particularly well suited for over-dispersed count data, such as ours, where the variance is greater than the mean. In the negative binomial model, the counts Y follow a Poisson distribution (λ), where λ is a random variable with a gamma distribution. Therefore, the unconditional distribution of Y is a negative binomial.