March 24, 2014

Institute of Population Health
University of Manchester
UK

Dear Sirs,

We would like the editors to consider our article entitled "ClinicalCodes: An online clinical codes repository to improve the validity and reproducibility of research using electronic medical records" for publication in PLoS One.

In this manuscript, we describe a new online database for lists of clinical codes (www.clinicalcodes.org) for use by researchers using electronic medical records (EMRs). This resource will allow for clinical researchers to better validate electronic medical records studies, build on previous clinical code lists and compare condition definitions across studies. It will also assist health informaticians in replicating database studies, tracking changes in disease definitions or clinical coding practice through time and sharing clinical code information across platforms and data sources as research objects.

Despite accurate definitions of medical conditions being a prerequisite for valid EMR studies and these definitions depending upon careful selection of clinical codes, the publication of clinical codes is rarely, if ever, a requirement for obtaining grants, validating protocols or publishing research. We evaluated the levels of transparency in the reporting of clinical code lists in a representative study of UK primary care database studies. Of the 392 studies we examined, only 35 (9%) published the entire set of clinical codes lists needed to reproduce or validate the study. These were most often published in online appendices.

We identify four main consequences of lack of transparency of clinical codes lists:

1. Code lists are not subject to scrutiny or peer review
2. It is impossible to tell if differences in found in study replications are genuine or due to artifactual differences in code lists
3. Comparisons between studies of the same clinical conditions are potentially invalidated
4. Lack of access to historical code lists leads to much wasted effort on the part of researchers

The database described here will provide a centralised repository for EMR researchers to deposit their codes and this will lead to greater transparency, reproducibility and validity in this important area of research.

We believe this submission fits all the PLoS ONE criteria for database papers, namely utility, validity and availability. The resource will be of great use to the EMR community and we expect the paper to be highly referenced and the ClinicalCodes database to becomes the de facto repository for clinical code lists across EMR research. The database is an effective repository for clinical code lists and we are aware no similar open repositories for clinical codes. The database is written entirely using open source software and is freely available for access, upload and download. In addition, we have developed open source software to access the database programmatically and to download research objects for integration with other systems.

We would like to recommend Irene Petersen from UCL as an Academic Editor.

Yours Faithfully,

David A Springate
Research Fellow
Institute of Population Health
University of Manchester