Reviewer’s report

Title: Integrating Historical Clinical and Financial Data for Pharmacological Research

Version: 4 Date: 27 September 2011

Reviewer: Pekka Ruotsalainen

Reviewer’s report:

The paper is now in good shape. Changes authors made are welcome and make the paper much easier to read and understand.

Some editorial comments. Correct mistakes in spelling and harmonize the way references are presented in the text, please. Correct also the way references are listed in the reference list.

Following examples can be extracted from page 8 s:

.. data scubbinng.[29] The ..
.. Wolters Kluver Medispan[31] (Wolters..)
.. metadata[26,27] attributes..

The correct way is to place references to the end of the sentence in the following way:

.. data scubbing [29]. The..

Reference should be presented in the "reference list" in the following way:

[1] Hersh W, The Electronic Medical Record..

There is something I like to mention for next steps: With EHRs, it is well known that all of its objects as codes, references and names have in real life errors up to 5-10%. This means that any code used as a truth code is still characterized by a statistical error. In the case of linking multi-source information with the help of codes, the cumulative impact of non-perfect linking codes can be mathematically expressed. It might to be valuable to present this error in a mathematical form.

Going back to this paper, I agree that the charge code (CC) seems be the most realible for record linking purposes.

Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:

I declare that I have no competing interests.