Author's response to reviews

Title: Integrating Historical Clinical and Financial Data for Pharmacological Research

Authors:

Vikrant G Deshmukh (Vik.Deshmukh@utah.edu)
N Brett Sower (brett.sower@hsc.utah.edu)
Cheri Y Hunter (cheri.hunter@hsc.utah.edu)
Joyce A Mitchell (joyce.mitchell@hsc.utah.edu)

Version: 4 Date: 7 September 2011

Author's response to reviews: see over
To

The Editorial Team,
BioMed Central

RE: Integrating Historical Clinical and Financial Data for Pharmacological Research
By Vikrant G. Deshmukh, N. Brett Sower, Cheri Y. Hunter and Joyce A. Mitchell
MS: 5319424745286299

Dear Editors,

We have revised our manuscript based on comments by the referees, and our proposed changes which address their specific concerns are outlined below.

We appreciate your thoughtful consideration of our manuscript for further review.

Sincerely,

Vikrant G. Deshmukh
N. Brett Sower
Cheri Y. Hunter
Joyce A. Mitchell

Referee 1:
The following major revisions were requested, and no additional discretionary revisions were recommended:

1. Method used for integration.
   a. The ‘vocabulary-based’ methods used for matching were essentially ontological in nature, and the same can be said of the ‘metadata-based’ methods that were developed as part of this work. It is important to make a distinction between the two methods compared in our manuscript, since the administrative metadata that were used for data-integration using metadata-based methods, were not true ontologies; consequently, the term ‘ontological-matching’ would only apply
loosely, and we have been careful to avoid using it to describe the overall matching approaches. Exact matches were required between the codes for controlled medical terminologies, vendor vocabularies and different types of codes for charges, dispenses, etc. that were stored in different systems, and these matches served as crosswalks between these systems. This part has now been clarified further on Page 9.

2. Practical matching example.

   a. We agree with the referee that a practical matching example would be helpful to the reader. The example for semantic enrichment of financial code, provided on Page 9 has now been enhanced with an additional figure (Figure 3), to illustrate it more clearly. A practical example was also added to the following section on integrating historical data on Page 10, and another figure (Figure 4) was added to illustrate the example of using RxNorm for integrating historical and current clinical data.

3. Use of acronyms.

   a. We agree with the referee that our manuscript contains a significant number of acronyms, and that some of them may be unique to the United States of America; however, the use fully-qualified names repeatedly in place of each of these acronyms would add to the length of the manuscript, and make it less readable. Since the BMC standard template did not include a separate section for abbreviations, a list of abbreviations has now been included as ‘Additional File 2’ in order to address the use of acronyms in our manuscript. We would welcome any recommendations that the editorial staff may be able to provide with regard to the inclusion of this list within the body of the manuscript itself.

**Referee 2:**

No major revisions were requested; the list below refers only to the discretionary revisions that were recommended:

1. Definition of Semantic Enrichment.

   a. We respectfully disagree with Referee 2, with regard to the appropriateness of describing these methods as ‘semantic enrichment.’ References to the use of this term in literature (Page 4, references 14-17) have been somewhat more inclusive in their scope and definition, with regard to the nature of information that the
enrichment processes added to the underlying data and metadata. The nature of additional information in the form of semantic relationships found in RxNorm that were added to financial data through methods described in our manuscript are consistent with the use of this term in literature.

2. Enhance the example on Page 9 with the addition of a figure.

   a. We agree with the referee that a practical matching example would be helpful to the reader. The example for semantic enrichment of financial code, provided on Page 9 has now been enhanced with an additional figure (Figure 3), to illustrate it more clearly. A practical example was also added to the following section on integrating historical data on Page 10, and another figure (Figure 4) was added to illustrate the example of using RxNorm for integrating historical and current clinical data.

3. How were name matches accomplished?

   a. The ‘vocabulary-based’ methods used for matching were essentially ontological in nature, and the same can be said of the ‘metadata-based’ methods that were developed as part of this work. It is important to make a distinction between the two methods compared in our manuscript, since the administrative metadata that were used for data-integration using metadata-based methods, were not true ontologies; consequently, the term ‘ontological-matching’ would only apply loosely, and we have been careful to avoid using it to describe the overall matching approaches. Exact matches were required between the codes for controlled medical terminologies, vendor vocabularies and different types of codes for charges, dispenses, etc. that were stored in different systems, and these matches served as crosswalks between these systems. This part has now been clarified further on Page 9.

4. General characteristics of the matching system used.

   a. The discussion on ‘Implications of the Findings’ has been expanded to include characteristics of the different systems on Page 22, which would make the work more easily reproducible by readers of the paper. An Additional figure (Figure 7) has also been added to further illustrate how these approaches could be generalized.