Author's response to reviews

Title: Accessing the public MIMIC-II intensive care relational database for clinical research

Authors:

Daniel J Scott (djscott@mit.edu)
Joon Lee (joonlee@mit.edu)
Ikaro Silva (ikaro@mit.edu)
Shinhyuk Park (shin_p@mit.edu)
George B Moody (george@mit.edu)
Leo A Celi (lceli@mit.edu)
Roger G Mark (rgmark@mit.edu)

Version: 2 Date: 19 November 2012

Author's response to reviews: see over
We would like to thank the Editor for this opportunity to revise and improve the manuscript. Please find our detailed responses to the reviewers below. All changes in the manuscript have been highlighted.

Reviewer 1

This paper introduces two tools that facilitate access to the MIMIC-II clinical data set. The MIMIC-II clinical data set is an extremely valuable publicly available large data set that not only provides a means for clinical research and research in the areas of clinical natural language processing, data mining and preservation of clinical research data, but also sets a precedent and an example for providing relatively easy and free access to clinical data not otherwise available to researchers. The tools enhance the value of the clinical data set by providing easy access for data exploration using QueryBuilder, as well as an opportunity to download and work with the data using a self-contained MIMIC-II VM. I have used both tools and had a positive experience with both, primarily getting answers to questions about data availability and potential cohort sizes for specific studies.

I am happy to see the paper that describes the data set and the tools. The paper also provides valuable pointers to additional information about the data, as well as the examples of clinical studies facilitated by the data and the tools. The paper is very well written and easy to follow. It will undoubtedly help inform researchers interested in clinical data and make the MIMIC data set more visible. I hope the paper will also make it easier for the other owners of clinical data to freely share their data sets with the research community.

We would like to thank the reviewer for the constructive feedback. Please note that all changes in the manuscript have been highlighted.

We also hope that the present paper will make MIMIC-II more visible among researchers and encourage data owners to share their data with the rest of the research community.

Minor issues not for publication:

The following discretionary revisions might make the paper even more useful for the researchers interested in accessing and using the MIMIC-II clinical data set and the tools:
Sections 3.1 and 3.2 clearly show the value of the MIMIC-II clinical data set. It would be good to also show the value of QueryBuilder and the MIMIC-II VM, for example, reiterate that the tools were used by the researchers to explore and download the data. It would be good to indicate when the variables were obtained using the VM and in what cases the researchers confirmed the availability of the data using QueryBuilder. Instead of "Using the data in MIMIC-II," you might explicitly state "Using the VM to obtain ..." Have the CinC 2012 participants used the tools or was the dataset generated using the tools?

For most MIMIC-II users, “using the data in MIMIC-II” is synonymous with “using QueryBuilder or the VM” because QueryBuilder and the VM are the most practical ways for accessing MIMIC-II data using relational database tools. The inseparability between MIMIC-II and the two access tools has been emphasized in the last paragraph of the Introduction (page 3) as well as in the first paragraph of the Discussion (page 9). We also mentioned in the Abstract (page 2) that QueryBuilder and the VM complement the value of MIMIC-II.

The example queries and studies in Sections 3.1 and 3.2 were conducted by us, using an internal connection to MIMIC-II. The PhysioNet/CinC Challenge participants did not have to extract the dataset themselves; the dataset was prepared by us. However, all examples in Sections 3.1 and 3.2 can be executed using either QueryBuilder or the VM. Immediately before Section 3.1 (page 6), we added our recommendation to use QueryBuilder for the simple example queries in Section 3.1 and the VM for the research studies in Section 3.2. Furthermore, we mentioned the demo VM (containing data from 4,000 expired patients in MIMIC-II) on page 5 as well as the maximum number of exportable rows in QueryBuilder on page 4, which are important considerations in deciding which tool to use.

If you are planning a survey or another user study (e.g., log files) to determine how the tools are used by the currently registered 300 users and the users’ preferences for one tool over the other, it would be good to mention these plans in the second paragraph of the discussion section. The last sentence of the discussion section mentions providing guidelines for formulating MIMIC-based studies. It would be good to summarize the guidelines in a list or bring the guideline items to the readers’ attention in the examples.

As of early November, the number of MIMIC-II users now stands at over 500, which has been updated on page 3. We provided some log-in and download statistics for QueryBuilder and the VM, respectively, in the second paragraph of the Results section (page 5). Also, we added a list of steps in the Discussion (page 9-10) as a guideline that a new user can follow in order to use QueryBuilder and the VM to conduct a clinical study.
Reviewer 2

The text is well written and very easy to understand.

My biggest problem with this work is on the research contributions. I view this paper as a very well written report on an engineering task consisting of adding query capabilities to the MIMIC II database using a J2EE server and also allowing users to access the database on a VM. These are all pretty standard engineering tasks and not really the product of extensive research.

We would like to thank the reviewer for the constructive feedback. We agree that both QueryBuilder and the MIMIC-II VM are results of routine engineering tasks, and we do not claim novelty from a software perspective. As noted in our reply to reviewer 1, however, QueryBuilder and the VM are the most practical ways for accessing MIMIC-II data using relational database tools. The inseparability between MIMIC-II and the two access tools has been emphasized in the last paragraph of the Introduction (page 3) as well as in the first paragraph of the Discussion (page 9). We also mentioned in the Abstract (page 2) that QueryBuilder and the VM complement the value of MIMIC-II.

We wish to position the present article as a description of two publicly available software tools that make it feasible for a broad community of researchers to perform innovative studies using MIMIC-II, thereby increasing its value. Moreover, to the best of our knowledge, we are not aware of other publicly available software tools that allow users to query a clinical database in SQL (mentioned on page 4). We also hope to ease a new user’s learning curve by providing example queries and studies as well as a step-by-step guideline (page 9-10) for using QueryBuilder and the VM to conduct clinical research.

The paper is complemented with case studies attempting to prove the usefulness of the system. It is unclear to me that QueryBuilder did play a significant role in these case studies. It appears to me that the studies were mainly successful because of the original MIMIC II database. I would urge the authors to address this point, may it be by running user studies measuring the efficacy of the proposed data access tools.

We would like to reiterate that QueryBuilder and the VM are the most practical methods to access the MIMIC-II relational database. Hence, any interested user would have to use either QueryBuilder or the VM to replicate the example queries or studies presented in the present article. It is true that the example studies were possible thanks to the rich clinical data contained in MIMIC-II, but practically speaking, it would be challenging for public users to conduct such studies if QueryBuilder and the VM did not exist.
We provided some log-in and download statistics for QueryBuilder and the VM, respectively, in the second paragraph in the Results section (page 5). The statistics show that both tools are routinely used by MIMIC-II users.