

Response to Reviews on

Improving the 'Fitness for Purpose' of Common Data Models through Realism Based Ontology
Jonathan Blaisure & Werner Ceusters

The paper proposes a RBO approach to improve the quality of CDM with OMOP CDM as an example. How to build a good data model structure is an important issue of the data model design.

The proposed RBO seems like a good solution but the reviewers would also like to see whether the authors can implement the work into the real database, and compare the performance between pre- and post-implementation, i.e. original OMOP vs. RBO, with an evaluation method mentioned in the method section.

It would be nice if the authors can provide the access to the proposed ontology model for readers. In general, this is a well written paper. However, some figures are suggested to be modified.

The reviewer also suggested that the authors can summarize the problems of current OMOP CDM using a table.

The content of the paper is suitable for AMIA conference. Please consider the comments suggested by reviewers for the final submission.

Reviewer's Comments

Reviewer #1

REV: The method for improving the quality of Common Data Models using Realism Based Ontology is well explained. The OMOP CDM was used to demonstrate the developed method. Problems were identified and possible solutions were proposed.

Authors presented a new way to look at the database structure as ontologists (provide a very high-level view of the world). I agreed with the authors' general claim: it is important for the ODM to accurately represent reality. But we need to simplify the world when we implement it.

The authors reported two types of problems in the OMOP CDM: (1) Confusion of types and (2) Reductionist problem involving temporality. Are there any other types of problems besides those two? It is better to have a table to summarize all types of problems.

AUTHORS: point well taken. Table 1 has been added as a summary to the paper.

REV: Page 5 the last sentence in the last paragraph: "This type of one-to-one relationship form the least costly joins that can be made in a relational database." But in the Figure 4, it showed as a one-to-many relationship.

AUTHORS: This figure was to show a possible solution to the problem, not the current state of the CDM. We addressed this by clarifying what each figure represents in the paper.

Reviewer #2

REV: The authors discuss about how a RBO approach can improve common data models in the sense of their fitness for purpose. They analyze some of the entities and attributes of the OMOP model and suggest how to reinterpret them based on selected ontologies.

The paper is well written and as the authors I see the advantages of using a RBO approach when designing data models. However, bringing it into practice is challenging and I do not know many real use cases where it has been done. I understand that what you present is on-going work. Are you planning to implement what you propose at your center?

AUTHORS: The IHI's primary function is to investigate and implement what we are calling a "Clinically Integrated Data Repository" (CIDR). As my work as a student is aligned with the RBO view, the work done here and ongoing work is to further that goal. We clarified our goal at the beginning of the "introduction" section.

REV: It would be very interesting to be able to compare the performance of the re-modelled CDM with the original OMOP and show real figures to the EHR modelling community.

AUTHORS: That would indeed be interesting, but we did not reach yet that phase in our work.

REV: I miss in the introduction mentioning other clinical information modelling standards such as ISO 13606, openEHR, HL7. Do you think they could be used as CDM? If not, why?

AUTHORS: We stated in the introduction the reasons why we selected OMOP. Since we only performed our analysis on OMOP, we can't answer that question on the basis of the work reported on. Nevertheless, HL7 and ISO 13606 as understood are specifications to enable interoperability between disparate health care systems. These specifications are useful in the process of populating a CDM and are currently being used (specifically HL7) at our institution. I agree that openEHR's database model may provide some insight into where some of the pitfalls lie in the road to creating a more robust RBO-based data model especially from a transactional point of view. A transactional database will certainly have a different "fitness for purpose" than a research based common data model.

REV: Many ontologies are supposed to be used in your reinterpretation, do you have publicly accessible somewhere the ontology model you produced? Adding to this work at least a link where the reinterpretation done is shown would be highly valuable.

AUTHORS: References to the ontologies have been added to the methods. We did not develop a new ontology but used the principles underlying the ontologies used to detect issues in the OMOP-model and to make suggestions how they can be corrected. We made it clear in the resubmitted paper.

REV: Figure 1 is not readable. Consider just adding a description of the complexity of the CDM instead. Improve the quality of Figures 3, 4 and 5.

AUTHORS: We agree. Figure 1 was removed and better resolution images added to the latter figures.

REV: Finally in the discussion I miss some lines describing why the authors think that such an approach has not been successful yet and how it could be improved.

AUTHORS: We are not aware of any similar approach, so we didn't have the suggested thought at all.

Reviewer #3

REV: In this paper, the authors present a realism based ontology approach to improve the quality of common data models. The authors explained the proposed approach using OMOP CDM as an example. It is important to design a good model to fit the purpose of database usage, not only for fast query, but also for efficient information retrieval.

The paper explained details of the methodology but it would be better to have some running samples to demonstrate the idea works, even if the sample size is small. The authors may consider to add the performance comparison between original OMOP CDM and RBO approach, and the evaluation in the method section. Otherwise it is not easy to make the conclusion that RBO approach really can improve the "fitness of purpose".

AUTHORS: It depends on what requirements drive the fitness for purpose. If a requirement is to have the least number of joins and the fastest latency, I agree RBO will not increase this particular fitness of purposes. A simple example would be an environmental exposure study that would require location history of all enrolled subjects. In the current state the CDM is incompatible with this purpose thus lacking. Though your point is well taken. There needs to be some metric to show the value in an RBO approach vs traditional design paradigms. Although performance is a valid metric we argue that the benefits of accurate reality based data outweighs performance requirements. With that said, from a practical stand point, we understand that the model has to be useable in the way that it cannot take months to query. As we continue our work, we will compare models using real world metrics.

REV: Since OMOP has wider purpose (as mentioned in the introduction), I'm also wondering if the single RBO approach is enough? Should we do more fine-tuning for different purposes? e.g. using different ontologies?

AUTHORS: The methods described in this paper are based on using the Basic Formal Ontology (BFO2) as a top-level ontology, but we have added also references to other BFO-compatible ontologies in the methods section such as the Information Artifact Ontology and the Ontology for General Medical Science. This can be expanded based on a wider purpose.

REV: In general, the paper is well-written and it introduces an interesting and important study, which is useful for designing a new data model. Yet the figure 1 may not give your work additional value but occupied half a page. I am looking forward to seeing the RBO approach can be implemented into OMOP CDM and having some interesting works with OHDSI.

AUTHORS: Figure 1 has been removed.