CONMEDS: Response to Reviewers

We thank the committee for their useful comments and suggestions. We have taken them all into consideration and substantially modified the plan to address reviewer concerns. The major issues were:

1. Potential pitfalls and a plan to address them (1,2,3,4,5,6)
For Aim 1, the matching of clinician-typed drug names will require manual validation; this is similar to a workflow in clinical trials called query management. We will manually review our dataset in its entirety to confirm the validity of the mappings and report the APIs' accuracy. We acknowledge that drugs recorded at enrollment time may not be taken during the trial itself; in order to maximize sensitivity we will leave these in. For Aim 2, a potential pitfall is incompleteness within the manually curated hierarchies. In this case, we will choose a relevant hierarchy that is most complete for our dataset. For Aim 3, although we have asked for academic permission to use the proprietary DrugBank side effect add-on, we have not yet received access. Our plan is to use accessible academic datasets, which are semi-curated and contain similar information.

2. Delineation of work, timelines, and environment (4)
Our group has domain expertise in clinical trials, and experience in developing user-friendly software tools. We are in contact with Stanford's Quantitative Sciences Unit, and received verbal approval to pilot this quality improvement initiative with its data from its director Manisha Desai. Karen is working on Aim 1, Bryan on Aim 2, and Holly on Aim 3. We have mostly completed Aim 1 as of April 30, and will have Aim 2 and 3 completed by May 30. We plan to connect the pipeline end-to-end by June 6.

3. Evaluation and metrics of success / choosing hierarchies (1,3,4,5,6)
For Aim 1, we will review non-exact matches. So far, only 17% of the drugs in our dataset have required approximate matching; for this subset, we will calculate the matches' accuracy through manual review. For Aim 2, if there are multiple candidate hierarchies, we will choose the hierarchy that is most complete for our dataset (as mentioned in 1) and discuss them with clinicians. For Aim 3, we will validate a random subset of patients' side effects and report accuracy.

4. Pipeline (4)
Our pipeline will be as follows: the user will input a CSV containing patient IDs and drug names. We will first call the RxNorm APIs to get the RxNorm ID and construct a new CSV. This updated CSV will be used to query the RxClass APIs for hierarchies, which will generate a list of parent nodes, and these nodes will be aggregated across our patient population to report summary statistics on their frequencies. This updated CSV will also be used to query the side effect APIs, which will generate a list of side effects and their associated frequencies. These will also be aggregated across our patient population to identify the most common side effects; they will also be reported on a patient-specific level to inform patient care. High-frequency hierarchy nodes and side effects will be highlighted in our clinician report in PDF form.

5. Lack of specifics beyond the API (extra comment)
We acknowledge that our proposal is more specific regarding the APIs, which are well-defined, while other aspects of the proposal are more vague. This is due to the fact that the hierarchies and databases we plan to explore in Aim 2 and 3 are not currently utilized within the clinical trial context. Our goal is to delve deeper into these ontologies to scope out their utility, and create a rank ordered list informed by characteristics including: data completeness, data quality, as well as data utility defined by the ontology's level of importance to overall patient safety or trial integrity.

6. Privacy (2) The data has been anonymized and there are no PHI concerns; we have replaced all patient names and the clinical trial identifiers with a new set of unique identifiers.