Summary of Menu-Driven Tool Testing for the ADAPTABLE Supplement

Heather Limper, Rebecca Johnston, Kimberly Barrett, Kathryn Goggins, Jessica Malenfant, Sunil Kripalani, Emily O'Brien

Background

Integration of patient-reported information and electronic health record (EHR)−derived data has the potential to both enhance evaluation of outcomes that are meaningful to patients and to improve data quality and validity. Patient-facing internet portals are increasingly utilized as an efficient mechanism for the collection of patient-reported information to answer key questions about disease manifestation, the patient experience, and other outcomes that are not readily available in EHR data. In the context of longitudinal studies, access to a patient portal may complement identification of clinical events by capturing outcomes that occur outside the patient's primary health system or after the patient relocates. Furthermore, patient web portals may facilitate clinical trial recruitment and enrollment through preliminary screening for eligibility and web-based consent. Despite these potential benefits, patient-reported information has not been systematically evaluated for completeness and validity, and key questions remain regarding its fitness-for-use in large-scale, pragmatic health research.

The ADAPTABLE trial, the first major randomized comparative effectiveness trial to be conducted by the National Patient-Centered Clinical Research Network (PCORnet),[1,2] aims to identify the optimal dose of aspirin therapy for secondary prevention in atherosclerotic cardiovascular disease.[3-5] As part of the new genre of patient-centered comparative effectiveness trials, the trial encompasses several key features, including enrollment of 15,000 patients across 40 health systems and health plans; an internet portal to consent patients and collect patient-reported information regarding risk factors, medications, healthcare utilization, and experiences; and reliance on existing EHR data sources for baseline characteristics and outcomes follow-up. Because ADAPTABLE relies on patients to report key information at baseline and throughout follow-up, it represents a unique opportunity to develop, pilot, and evaluate methods to validate and integrate patient-reported information with data obtained from the EHR. The objective of this study was to develop methods to 1) assess the quality of patient-reported data and 2) to integrate the data with existing electronic health data. To achieve that end, this study assessed the concordance of patient-reported hospitalizations with those documented in the EHR.

Methods

Menu-driven queries (MDQs) utilize a simple point-and-click interface to add terms to the request criteria that are joined through logical operations (and/or), associations, and then grouped according to the chosen stratification(s). A major benefit is that users do not need programming expertise to create and execute MDQs.

The ADAPTABLE Supplement project developed an MDQ to enable rapid comparison of patient-reported health data to analogous data in the EHR to determine concordance between the two sources for hospitalizations occurring during the trial follow-up period. This MDQ was based on the PCORnet Common Data Model (CDM) to enable broad uptake by participating sites across the network. Functionality was developed to include the ability to customize date ranges, outcomes of interest, trials that participate in the CDM, and a variety of other parameters.

The tool was tested using data from a population of patients enrolled in the ADAPTABLE trial by Vanderbilt University Medical Center (VUMC), which is part of the Mid-South Clinical Data Research Network and was the top enrolling site in the U.S.[6] We used the MDQ to compare the concordance of HER-recorded hospitalization events and those directly reported by patients via an online study portal. At 3-6 month intervals, patients enrolled in this trial were asked to provide any recent hospitalization events including the date, reason, and location of the hospitalization. Location (hospital name, city, and state) was entered as free text. Date was entered into formatted fields for month, day, and year. Reason was selected from 7 check boxes relevant to the

study outcomes (chest pain, heart attack, angioplasty, cardiac bypass surgery, transient ischemic attack-like symptoms, stroke, or major bleeding), or marked as none of the above.

When the data were queried from the patient portal, a single hospitalization was associated with 3 rows: 1) the reason for hospitalization (HOSPITALIZATION_EVENT), 2) the location of the hospital encounter (HOSPITALIZATION_LOCATION), and 3) the date of the event (HOSPITALIZATION_DATE). This patient-reported data were routinely uploaded into the CDM during the larger ADAPTABLE trial, which allowed for linkage to data available in the EHR within the VUMC DataMart. The MDQ was run and tested against a use case chosen by the study team (described below).

Use Case

The MDQ was executed to query: *"For patients in the ADAPTABLE trial, find all EHR-based encounters occurring within 7 days of a patient-reported hospitalization, as well as all of reported hospitalization events for those patients."* Hospital-based encounters were only returned if a patient had both: ≥1 patient-reported hospitalization date that had a matching EHR encounter within 7 days AND ≥1 patient-reported hospitalization event. The use case set out to compare the dates of hospitalization reported by patients to those documented in the EHR. An alternate version of this use case was also attempted using a +/- 21 day window, the time frame being used for outcome adjudication in the ADAPTABLE trial.

Output

The MDQ was developed to return EHR-derived outcomes and patient-reported outcomes in the same spreadsheet to facilitate review, providing a single row per matched hospital encounter and date within a 7-day window (Figure 1). Data were returned for each matched hospital-based encounter with one row per encounter-diagnosis pairing.

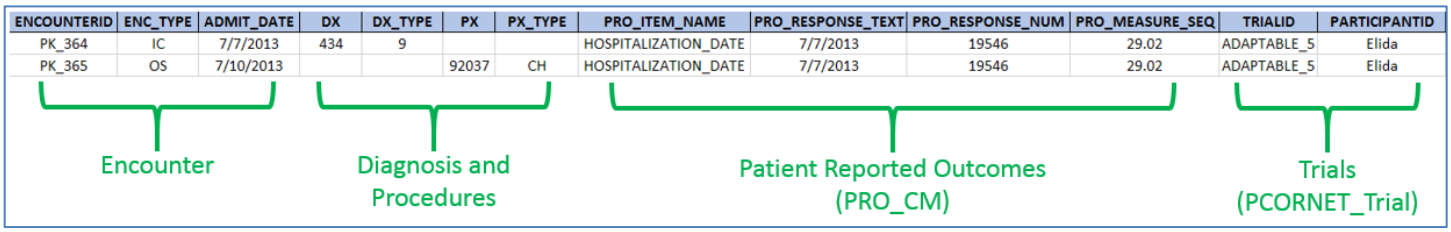

Figure 1. Output from the Menu-Driven Query

Results

The developed MDQ was run against data from May 20, 2016 through May 31, 2019. A total of 33,985 rows were returned by the query, representing patients with a patient-reported hospitalization event who had at least one encounter documented in the EHR within 7 days of a patient-reported hospitalization date. These rows were attributed to 399 unique study participants, which represents 58% of participants who entered patient-reported information in the ADAPTABLE study portal. Patients who were excluded from the query include those for which patient-reported date and date in the EHR were more than 7 days apart as well as patients who provided either a hospitalization event or event date but not both in the portal.

TABLE 1. Days of concordance

Days	Count
-7	845
-6	936
-5	705
-4	661
-3	1359
-2	1271
-1	3207
0	14555
1	2242
2	969
3	1020
4	947
5	941
6	840
7	1327
Total	31825

Among the 33,985 rows, each of which represents a pairing of a hospitalization event and a diagnosis, 45.7% of EHR-documented dates were an exact match with patient-reported dates. A total of 84.6% of EHR-documented dates were within 5 days of patient-reported dates (Table 1). Encounters were categorized in the CDM into 4 categories: ambulatory visit, emergency department (ED), inpatient hospital stay, or no information. Ambulatory visits were most accurately reported with an average -0.08 days separating patient-reported and EHR-documented date of encounter (Table 2). ED visits had the least variation in concordance between patient-reported and EHR

documented date of encounter with a range of -1 to 6 days separating these dates. Of note, ED visits include ED encounters that became an inpatient stay.

Table 2. Reporting accuracy by encounter type

Encounter type	Count	Average days	Min days	Max days
Ambulatory visit	17845	-0.08	-7.00	7.00
Emergency department	384	0.17	-1.00	6.00
Inpatient hospital stay	15656	0.26	-7.00	7.00
No information	100	0.52	-7.00	7.00
Total	33985	0.08	-7.00	7.00

When the alternate 21-day version of the query was attempted, the query ran for more than 19 hours, at which point it timed out due to security parameters set on the VPN used to access the CDM. Thus, the 21-day query could not be executed despite multiple attempts.

Exact matching on HOSPITALIZATION_LOCATION was not performed due to a high degree of variation in how hospital names were typed in the patient-reported data, including variations in name, spelling, etc. For example, one facility (Vanderbilt University Medical Center) was recorded 50 different ways by patients, with variations including "Vanderbilt", "Vanderbilt University", "Vanderbilt University " (with an extra space at the end), "VanderbiltUniversity Medical Center" (missing a space), "Vanderbilt University MC" (abbreviated), "Vanderbuilt" (misspelled), "Vanderbilt University Hospital", "VUMC", and so on. Similar permutations were observed with other facility names (e.g., "Maury Regional Medical Center" was typed 19 different ways, and "Williamson Medical Center" was typed 21 different ways). Among the 399 unique patients, there were exactly 398 different hospital names typed, indicating that only once did two patients type out the hospital name in the same way. Therefore, hospital location was not considered in determination of concordance.

Discussion

An MDQ was deployed, proving feasibility of this tool as a method for querying widely available data sources such as the PCORnet CDM. With 45.7% of EHR-documented dates being an exact match with patient-reported dates and 84.6% falling within 5 days of reported encounters, results indicate relative reliability of patient-reported and EHR-based event dates. However, matching for location and for longer time windows proved more problematic.

A few limitations were noted. First, the matching procedure required an exact match, and while this was straightforward for hospitalization date, it was much more complex for hospitalization location due to variations in the way patients spelled or typed the hospital name. Manual review or a fuzzy match would be required to match location when free-text patient data entry is allowed. Second, the PRO_CM table structure inhibits the ability to have one row for all information associated with one hospitalization. The MDQ generated multiple rows for each event, connected by a linking variable (PRO_MEASURE_SEQ), and data manipulation was required in order to assess concordance. Third, as noted above, a query with a broader time range of +/- 21 days required a very long run time and timed out under the cybersecurity parameters set for the VPN. To resolve this would require changing the institution's security parameters, working with a random sampling of the data, or restructuring the data in a more efficient format, which were not feasible under the scope of this project.

Conclusions

We found that use of an MDQ tool to examine concordance between patient-reported and CDM data were feasible but had several limitations in its execution, including the time window used to search for concordant events. This project has the potential to inform future efforts to synthesize potentially inconsistent data from patient-reported and EHR sources, to identify opportunities to streamline data capture, and to facilitate enrollment of study-specific target populations within larger health systems.

References

1. Collins FS, Hudson KL, Briggs JP, Lauer MS. PCORnet: turning a dream into reality. *J Am Med Inform Assoc.* 2014;21(4):576-577.
2. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc.* 2014;21(4):578-582.
3. Jones WS, Roe MT, Antman EM, et al. The changing landscape of randomized clinical trials in cardiovascular disease. *J Am Coll Cardiol.* 2016;68(17):1898-1907.
4. Johnston A, Jones WS, Hernandez AF. The ADAPTABLE trial and aspirin dosing in secondary prevention for patients with coronary artery disease. *Current cardiology reports.* 2016;18(8):81.
5. Hernandez AF, Fleurence RL, Rothman RL. The ADAPTABLE Trial and PCORnet: shining light on a new research paradigm. *Ann Intern Med.* 2015;163(8):635-636.
6. Rosenbloom ST, Harris P, Pulley J, et al. The Mid-South Clinical Data Research Network. *J Am Med Inform Assoc.* 2014;21(4):627-632.