Supplementary Table S1 Questions assessing technical capabilities of sites during site selection interviews

Site EMR Background
1. After reading the Study purpose and design and reviewing the list of data elements, what are the biggest concerns that you have about participating in this Study?
2. If your Site participated, would you be able to designate a primary technical contact acting as an IT/Clinical data Subject Matter Expert (SME) and Liaison throughout the life of the study? Backup person available?
3. Is your institution now, or planning to be, a site/member in the PCORnet Network? If so, how far along are you with mapping, testing, etc.?
4. Have you previously extracted data from your Institution's EMR for other studies or for government agencies? Describe your largest project involving EMR data extraction for an external entity.
5. Which EMR system(s) are you using? (i.e., Epic, McKesson, Homegrown, etc.)
Sites Governance/Rules
1. What constraints exist within your institution for exporting data outside of your institution (e.g., firewalls)?
Data Elements Review/Analysis
2. What % of the data elements from the list provided by DCRI, do you estimate you can provide?
3. After reviewing the data elements, you may find that your data do not exactly match the study requirements, especially the value sets/enumerations. Will you be able to share your Data Dictionary of all of your data points that include your value sets and definitions?
4. Which standards does your EMR system use for identifying conditions, diagnoses (ICD9, ICD10, SMOMED, other), and procedures (CPT, LOINC, NCD, RxNorm, other)? Do you use your own in-house (local) code lists to identify items listed in the Data Elements list?
5. Regarding PHI, will you be able to provide deidentified datasets?
6. If Yes for the above question, will you be able to track your real Patient identifiers to surrogate identifiers that are meaningless except to you? This requires that if a patient has >1 admission that the datasets provided by you would have the same surrogate patient identifier for the multiple admissions. We will still need limited PHI data elements like patient date of birth, dates of admission and discharge, dates of diagnoses and procedures
7. What types of discrete data can you obtain from Flowsheets used for inpatients?
8. Do you track if your inpatients are enrolled into other Studies?
Data Extract & Transfer:
1. Would you be providing a full extract of all admissions occurring from the beginning of the study period, or an incremental file containing admissions done since the last extract occurred?
2. What are your standard extract formats (e.g., xml, SAS, HL7 CCDA, spreadsheets)?
3. Data transfers containing PHI will likely require encryption (not just SFTP). Any experience with using public encryption keys to encrypt files before you transfer?
4. How do you prefer to do the data transfer? We can pick up files within your firewall or you can push to our location.
5. How much of the data are already collected in a data warehouse or ODS?
6. How are data quality checks performed before data are introduced to the EDW?
7. Do you have any additional questions or comments for DCRI?