

will meet ALCOA standards in their present format.

The extent of the problem collecting data from disparate systems is obvious in the following example from Sweden; there are eight quality registries using thirty five individual technical platforms and there is no interoperability or standardization between systems, which prevents valuable analysis via data pooling.

By contrast, the Netherlands has a single platform strategy so there is conformity in the way the data is ingested, stored and standardized. In addition there is a three year process in place to allow each provider time to ensure that their data meets the quality standards before data is considered for reporting.

Ideally, using data from ONC certified technology is preferable, however, this is a US initiative and it is more likely that data will be solely mined from EHR's which are not certified by ONC and will require understanding of the source data and defined clinical data management processes to ensure the validity, reliability and integrity of the data. The eClinical Forum's eSource Readiness Assessment (eSRA) is a very useful tool, which provides the minimum requirements for self-assessment for systems using data that may be included in a clinical trial. Going back to the fundamentals of ALCOA, sponsors will have to ensure that system access is limited to authorised users, those authors are identifiable, there is an audit trail and records are available for inspection and the privacy and security of patient data is safe guarded.

The future of using EHR and other real world data lies in the process, semantic (ability to understand the data), and technical interoperability

between systems supporting clinical trials following ALCOA principles to ensure the integrity, validity, and reliability of the data that has been mapped and extracted or exchanged. Integration activities must also comply with legal and regulatory requirements to protect patient's privacy and security of data. This in itself is a serious undertaking for any organization.



**Clinical trials are designed to answer specific questions and the CDM process is designed to deliver valid and reliable data for statistical analysis**



Our experience to date collecting data from EHRs/patient reported systems indicates you will need to develop a robust version controlled data dictionary, which clearly defines each data item you wish to collect, supporting definitions, response options, the timing of when data was collected and the reporting source. We have also noted that using coding standards such as ICD10 for missing data may improve the quality of the data. You will need to have a clear understanding of how the data is collected in the source systems you are accessing/ accepting data from. Nuances in the clinical practices introduce variability, so you must allow for this in your data transfer specification/ mapping tool. This means spending time speaking to the data managers at each provider to ensure that the data extracted meets the definition of the data dictionary. A gap analysis

and pilot testing helps to identify the degree of missing data and indicates where there are issues with validity. Structural checks should be run over data received from each provider to ensure that the data extracted meets the data specification. This will not however tell you that they mapped the data item correctly and even with the best instructions, people will have their own interpretations. Therefore a process or data audits will need to be implemented.

Classic data cleaning will no longer be applicable as the data you receive is source data. Unless the data provider is able to correct data within the EHR system after receiving feedback about the data issues, the expectation is that missing or odd values will be carried through the process. It will then be a clinical and statistical decision as to whether records are included in the analysis or not.

Notably the integrity of some data points can only be checked with visual review, for example, Patient Reported Outcome instruments need to have been migrated faithfully using the ePRO consortiums best practice guidelines. This implies the need to access the patient facing materials and to assess that the instrument has not been changed.

Another challenge is the variability of languages. Case Report Form data is generally captured in English language. The EHR systems usually capture data in local language and therefore this needs to be taken into consideration.

This is a new era for clinical data management and new processes and methods will be necessary to manage this new data source to ensure the validity, reliability and integrity of the data, as there is little point in collecting data if it cannot be trusted to meet the quality standards reliably. 