Royal Papworth Hospital
NHS Foundation Trust

GD014

Clinical Data Management Validation

1. Validation is completed to give assurance that an identified process will meet pre-set specifications and quality characteristics. The term validation in Clinical Data Management is usually used to describe different aspects of electronic data. This document should help in the understanding what validation refers to by explaining the four different meanings, which are:
    a. Computer System Validation
    b. Build Validation
    c. Data Validation
    d. Source Data Verification
2. Computer System Validation
    a. This is the validation of computer systems and software, or data repositories that house the data, making sure these function as expected.
    b. This is covered by section 14.5 of the Good Clinical Practice Guide (2012)
    c. Commercial Systems
        i. Commercially produced and validated software, often referred to as “off the shelf”, for example, MS Excel, SPSS and STATA are validated by the software developers before being released for sale, as well as a multitude of users, and should be used wherever possible.
        ii. These should be validated as fit for purpose, risk assessed with due diligence. If there are manufacturer tools for validation, these should not be the only validation documents, and can only be used if they are documented in the service agreement.
        iii. When commercial software is updated it is suggested that the software is tested before use. As many software packages are updated frequently and sometimes without the users knowledge, only major upgrades are covered here. For example Windows 7 to Windows 8, or OpenClinica 3.6 to 3.7. For large scale software packages, (for example MS Excel) this might only involve only comparing the files in the two versions. For smaller scale software, for example OpenClinica, running checks on the new version, preferably before making the upgrade widely available and documenting these tests is best practice.
        iv. OC Validation – These are stored in S:\shared\OpenClinica\Documentation
            1. Installation Qualification (IQ)
            2. Operational Qualification (OQ)
            3. Performance Qualification (PQ)
            4. Upgrade validation.
    d. Bespoke Systems
        i. For Papworth Sponsored CTIMPs, bespoke data collection software cannot be used without a detailed explanation in the Data Management Plan (DMP) and approval by both the Clinical Project Manager and the Data Manager. This is because the validation requirements are complex.

   ii. For non-CTIMP studies, bespoke data collection software should also be avoided, but if a decision is made to use such a system this should be documented in the DMP.
   iii. Bespoke programming can be used for other tools that are not data collection, but these must be validated.
   iv. Validation for bespoke systems must include fully documented User Acceptance Testing (UAT) and risk assessments. The UAT should be test every part of the system, and extend beyond the data validation plan based UAT.

3. Build Validation (The UAT guidance document, GD008 expands on this.)
   a. This is the validation of the study data tool which include the data repository and the Case Report forms (CRFs) (paper and electronic).
   b. As a database/eCRF is developed, it should be continually tested.
   c. While testing is essential in the build of databases/eCRFs, the paper CRFs should also be tested before use. This should include reviewing the order of questions, the units requested, the space/format of the answer and the usability.
   d. UAT is a required validation step for database/eCRFs, see GD008 for further documentation.
   e. For CTIMPs scripts for testing should be based on the DVP if available. It is preferable that multiple user types run the test, from the data team, to the PI or CTC, to members of staff naïve to the software and/or the study. This all should be documented, and repeated after each change in the design, until there are no errors to resolve.
   f. For non-CTIMP studies it is possible to fast track the UAT, which would involve a shorter script.
4. Data Validation (The cleaning guidance document, GD012 expands on this.)
   a. This is the validation of the actual data by designing methods to locate errors.
   b. This is covered by section 8.5 of the Good Clinical Practice Guide (2012)
   c. The following are examples of data validation:
      i. On Entry Validation
         1. Double data entry
         2. Ranges
         3. Required fields
         4. Finite options for responses (e.g. drop down lists)
         5. Expected formats
         6. Logic checking data (e.g. males are not pregnant)
      ii. Post entry checking
         1. Automated Edit Checks
         2. Reconciliation (often for SAEs)
5. Source Data Verification (SDV) (SOP016 Monitoring Research Studies expands on this.)
   a. This is the validation that the raw/source data matches data in the data repository.
   b. While SDV is verification rather than validation, this is covered in this guidance document as the term validation sometimes is used when SDV is meant.
   c. SVD is the process by which data within the case report form (CRF) or other data collection systems are compared to the original source of information (and vice versa).