Royal Papworth Hospital
NHS Foundation Trust

GD014

Clinical Data Management Validation

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 understanding what validation refers to by explaining the four different meanings, which are:

1. Computer System Validation
2. Database Build Validation
3. Data Validation
4. Source Data Verification

1. 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) (MHRA Grey Book).

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 and risk assessed with due diligence. Manufacturer tools for validation may be used only if they are documented in the service agreement. However, they should not be the sole validation documents.

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 user's knowledge, this section will focus on updates where users are informed. For instance, large scale commercial systems, like Windows 10 to Windows 11, or MS Office 2016 to MS office 2021 might only involve comparing the patch release notes between the two versions.

iv. For commercially smaller scale software like OpenClinica, running and documenting validation checks on the new version, preferably before making the upgrade widely available is best practice. The validation reports and documentations for all OpenClinica upgrades at the Papworth Trials Unit Collaboration (PTUC) can be found in S:\shared\OpenClinica\Documentation

- Installation Qualification (IQ)
- Operational Qualification (OQ)
- Performance Qualification (PQ)
- Upgrade Validation.

v. See GD026 Data Management Software Management for more information on OpenClinica upgrade validations.

d. Bespoke Systems – The technical details of these systems go beyond the scope of this guidance document, but they involve building the entire system from smaller components, typically handled by specialist programmers or developers.

i. For Papworth Sponsored CTIMPs, bespoke data collection software must not be used without a detailed explanation in the Data Management Plan (DMP) and approval by the sponsor. 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 test every part of the system and can extend beyond the validation checks in the UAT document.

2. **Database 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, UAT is primarily conducted using detailed test scripts that ensure all relevant section of the study database are validated. In addition to the data team, it is preferable to have different user types run the UAT. For example, the Clinical Data Lead (CDL), Clinical Trial Coordinator (CTC), or members of staff naïve to the software and/or the study. This should all 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 (see GD008 – UAT for more guidance).

3. **Data Validation** (refer to GD012 Data cleaning for more guidance)
    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)

4. **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. SDV 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).
    c. 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.

Referenced Documents
SOP016 Monitoring Research Studies
GD008 User Acceptance Testing
GD012 Data Cleaning
GD026 Data Management Software Management