FUNDED BY NIHR | National Institute for Health Research

UKCRF Network Quality Assurance Theme Group

Advisory Notes for Data Integrity

Version 01, October 2019

Contents

1. Introduction 2
2. General Concepts and Principles 3
   2.1 Data Life Cycle 3
   2.2 Risk Management 4
   2.3 Data and Metadata 5
   2.4 ALCOA(C) Principles 6
3. Source Data 7
4. Computerised Systems 10
   4.1 User Access and Permissions 10
   4.2 Security and Backup 10
   4.3 Electronic Approvals and Signatures 10
   4.4 Computer Systems Validation 11
5. Audit Trails 12
6. Electronic Health Records (eHRs) 14
7. Electronic Patient Reported Outcomes (ePROs) 17
8. References and Resources 19

1. Introduction

Ensuring the credibility of clinical trial data has always been one of the key objectives of Good Clinical Practice (GCP). Decisions made on the back of study outcomes can only ever be as good as the quality of the data which support those outcomes. Research data must therefore be of sufficient quality to support good decision making, and furthermore, there must be evidence of this quality to instil confidence in the decisions made.

> *"All clinical trial information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation and verification. This principle applies to all records referenced in this guideline, irrespective of the type of media used."*
>
> ICH E6(R2) Guideline for Good Clinical Practice

As part of their quality system, Clinical Research Facilities (CRFs) should have processes in place to maintain data integrity. However, the increasing use of electronic systems in modern clinical research has presented some new challenges.

Data integrity has been a hot topic and a focus of regulatory scrutiny in recent years. As this scrutiny has grown, so has the amount of guidance on the subject from multiple regulatory and industry sources. In March 2018, the Medicines and Healthcare products Regulatory Agency (MHRA) issued a 'GXP' Data Integrity Guidance and Definitions document, which aims to facilitate improved compliance by clarifying the regulatory interpretation of the requirements in the UK. In this document, data integrity is defined as follows:

> *"Data integrity is the degree to which data are complete, consistent, accurate, trustworthy, reliable and that these characteristics of the data are maintained throughout the data life cycle."*
>
> MHRA 'GXP' Data Integrity Guidance and Definitions, Mar 2018

The challenge for CRFs is to ensure that study data fulfils these attributes to a satisfactory degree. Whatever the format, paper or electronic, it is important that CRFs are able to demonstrate that the data they are generating, processing, handling or storing is suitably reliable and secure. However, deciphering the mass of recent guidance and understanding the implications for CRFs is not always straightforward, particularly when the concepts and terminology used in an increasingly digital era may be largely unfamiliar.

The purpose of these advisory notes is not to rewrite guidance that has previously been issued elsewhere. Neither should they be considered a complete and definitive reference point, encapsulating the entirety of the requirements for CRFs. For this reason, relevant references are signposted throughout and should be consulted for further details and more in-depth guidance on regulatory expectations. However, it is hoped that this document will help to explain some of the jargon, improve understanding and give some simple practical advice for ensuring data integrity in a CRF.

2. General Concepts and Principles

2.1 Data Life Cycle

It is important to recognise that there are many points where risks to the integrity of study data might exist. This is clear from the MHRA definition of data integrity, which states that this must be maintained "**throughout the data life cycle**".

The data life cycle is simply the different phases that the data will go through, from the moment it is generated until it is destroyed. Although the specific details will depend on the data, the typical phases of a data life cycle are shown in Figure 1.

The extent of CRF interaction with study data will vary with the nature of the study and the data. CRFs should be clear what stages of the data life cycle they are responsible for or can directly or indirectly influence.

For example:

- Entries in an electronic patient diary may be transmitted directly to a central database, with no opportunity for CRF manipulation. However, CRF staff may be responsible for training patients in the use of the system.
- A patient's BMI may be recorded and retained by the CRF, but there may be no involvement in archiving or destruction of the data.
- For temperature data from a freezer used to store samples, the CRF may be responsible for every stage of the data life cycle.

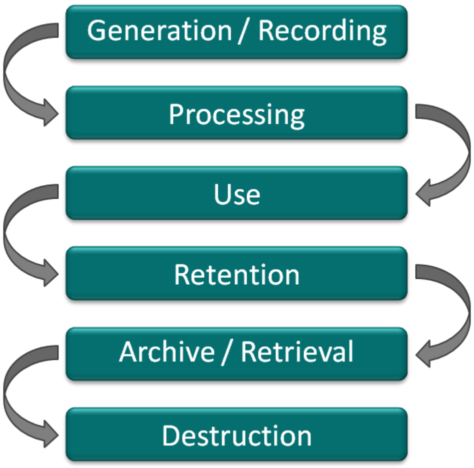


Figure 1. Phases of a data life cycle.

Each phase of the data life cycle and the associated risks must be considered. Focussing attention on one phase while neglecting others will not ensure data remains credible - a data integrity breach can be equally devastating, no matter what phase it occurs in. Efforts taken to ensure data is generated, recorded and processed accurately may be wasted if it is not stored in such a way as to protect it from unauthorised manipulation or deletion.

Ask yourself...

- *What study data does the CRF interact with? What phases of the data life cycle is the CRF responsible for or can directly or indirectly influence?*
- *Have controls been put in place to ensure data integrity is maintained throughout all phases of the data life cycle that involve the CRF?*

2.2 Risk Management

Quality can be defined as the absence of errors that matter. Data integrity is not about ensuring perfect data, but rather eliminating or mitigating any data risks that have the potential to jeopardise participants or to undermine confidence in study outcomes and decisions. CRFs should therefore focus the greatest effort and resource on ensuring the quality and reliability of critical data, i.e. the data that really matters.

One way to do this is to carry out a data integrity risk assessment. This can be done in different ways, but some of the key aspects to take into consideration are shown in Figure 2. Risk mitigation strategies should be realistic and commensurate with the data criticality. They should take into account an acceptable level of residual risk, as reducing the risk to zero is often impractical or even impossible.

Further guidance on risk assessment is given in the MHRA 'GXP' Data Integrity Guidance and Definitions document.

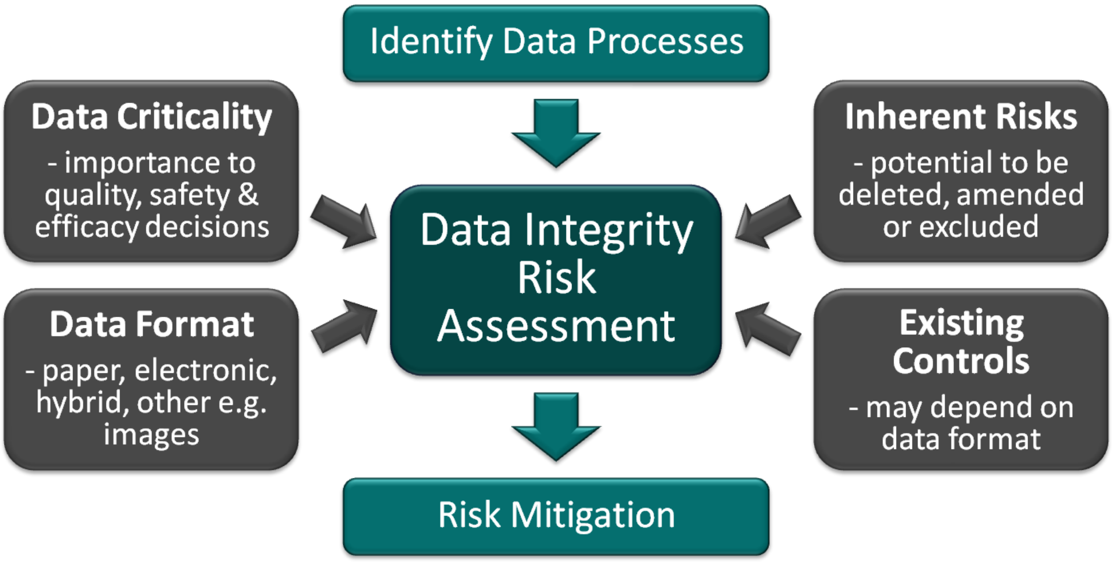

Identify Data Processes
Data Criticality
- importance to quality, safety & efficacy decisions
Inherent Risks
- potential to be deleted, amended or excluded
Data Integrity Risk Assessment
Data Format
- paper, electronic, hybrid, other e.g. images
Existing Controls
- may depend on data format
Risk Mitigation

Figure 2. Considerations for a data integrity risk assessment.

Ask yourself...

- *What study data does the CRF interact with that is critical to participant safety or study outcomes? Do processes associated with the data pose a risk of a data integrity breach?*
- *What risk mitigation steps can reasonably be taken to ensure the security and credibility of the data? Is the remaining residual risk acceptable?*

2.3 Data and Metadata

Clinical trial data is any information, whatever the format, which allows reconstruction and evaluation of trial activity. However in most cases the data itself is of limited use in isolation, and requires supporting information to allow it to be interpreted and evaluated. This "data about data" is often referred to as the **metadata**, and can include such aspects as participant identifiers, units of measure, time points, who generated the data, etc.

> *"Metadata are data that describe the attributes of other data and provide context and meaning. Typically, these are data that describe the structure, data elements, inter-relationships and other characteristics of data e.g. audit trails. Metadata also permit data to be attributable to an individual (or if automatically generated, to the original data source)."*
>
> MHRA 'GXP' Data Integrity Guidance and Definitions, Mar 2018

Metadata is equally as important as the data it supports, as without it the data has no meaning. It should therefore be subject to the same level of protection and control. Care should be taken to ensure that appropriate metadata is captured and maintained throughout the retention period of the data.

With purely paper records this is reasonably straightforward as the metadata, such as dates and signatures, is usually entered as a permanent record along with the data. It therefore remains linked to the data for as long as the paper record (or a certified copy) is retained.

However, data generated or stored in an electronic format may require further consideration. This may simply be because the electronic system does not record all the required metadata. For example, a printout from an ECG or BP monitor may need to be annotated with identifiers, dates, time points, etc. Alternatively, the system may capture the metadata but it does not remain linked to the data it supports due to how the data is extracted and stored. This can be a particular issue when converting a "dynamic" file to a "flat" file, e.g. printing a page from an electronic system, or converting it to a PDF.

Retention of metadata is important during the trial, but also throughout the archive period. It is here that metadata is often lost, as it has not been taken into consideration when making arrangements for how the data will be archived. For example, if the metadata is held in a separate but dynamically linked database within an electronic system, this can be lost if only the data file is exported and archived. This is a good example of a failure to consider the risks to data integrity throughout the whole data life cycle.

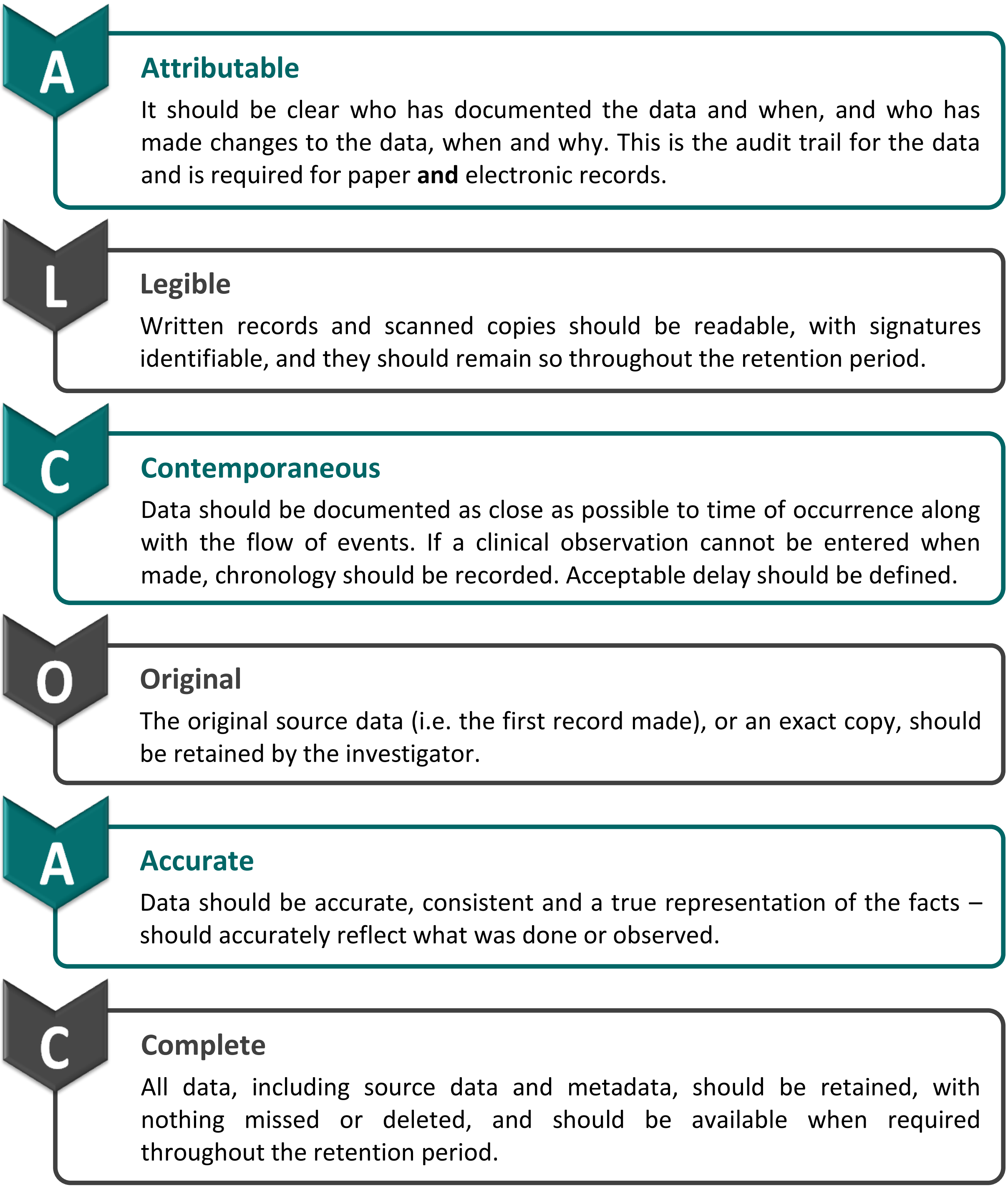
Ask yourself...

- *What systems, paper or electronic, are used to capture study data in the CRF? Do they capture ALL the required metadata? If not, how else could this be recorded?*
- *How is study data extracted from electronic systems in the CRF? Does associated metadata remain linked to the data? Is this true throughout the retention period, including archiving?*

2.4 ALCOA(C) Principles

ICH E6(R2) Guideline for Good Clinical Practice states that source documents and trial records should be **attributable**, **legible**, **contemporaneous**, **original**, **accurate**, and **complete**. These key attributes are sometimes referred to as ALCOA(C) principles, and are applicable to any clinical trial data entries or changes. Some general considerations for applying these principles are outlined in Figure 3.

Figure 3. Considerations for applying ALCOA(C) principles to trial data.



Attributable

It should be clear who has documented the data and when, and who has made changes to the data, when and why. This is the audit trail for the data and is required for paper **and** electronic records.

Legible

Written records and scanned copies should be readable, with signatures identifiable, and they should remain so throughout the retention period.

Contemporaneous

Data should be documented as close as possible to time of occurrence along with the flow of events. If a clinical observation cannot be entered when made, chronology should be recorded. Acceptable delay should be defined.

Original

The original source data (i.e. the first record made), or an exact copy, should be retained by the investigator.

Accurate

Data should be accurate, consistent and a true representation of the facts – should accurately reflect what was done or observed.

Complete

All data, including source data and metadata, should be retained, with nothing missed or deleted, and should be available when required throughout the retention period.

3. Source Data

Adhering to ALCOA(C) principles is important for any trial documentation, but no more so than when it comes to source data. Source data can be thought of as the original record for any trial information. In other words, it is the very first capture of the information, whether that is in paper, electronic or any other format. A definition is given below.

> *"All information in **original** records and **certified copies** of original records of clinical findings, observations, or other activities in a clinical trial necessary for the reconstruction and evaluation of the trial. Source data are contained in source documents (original records or certified copies)."*
>
> ICH E6(R2) Guideline for Good Clinical Practice

It is not always straightforward to determine what the original record is, and in some cases this may not be in a format that can be captured easily. However, it is generally accepted that the earliest practically retainable record should be considered the source. For example:

- Where data is generated in electronic format, but the instrument used does not store the electronic data (e.g. BP monitor or ECG), a paper printout from the equipment can become the source.

- If the equipment does retain the electronic data but only stores it temporarily, it may be possible to periodically extract the data or reconcile it against a paper record for longer term retention as the source.

- In situations where a participant transmits data via an app or wearable device to a central server, and the device itself does not retain the data, then the data on the central server would be the source.

If the original record is not retained, then it may be possible to make a "certified copy" which can be considered the source. A certified copy is defined as:

> *"A paper or electronic copy of the original record that has been verified (e.g., by a dated signature) or has been generated through a validated process to produce an exact copy having all of the same attributes and information as the original."*
>
> ICH E6(R2) Guideline for Good Clinical Practice

Care must be taken to ensure no information is lost during copying. For example, common issues with scanning include double-sided documents being scanned single-sided, information close to the edge of pages being cropped off, pages being folded over when scanning or the scan resolution being so poor that e.g. handwriting is no longer legible.

Importantly, "all of the same attributes and information" includes any associated metadata. So if a printout from an electronic system does not include the audit trail data, or a document containing important colour information is scanned in black and white, then these cannot be considered certified copies.

Source documentation serves as the medical record of the participant and should document their progress throughout the trial. This includes confirmation of eligibility, the consent process, study visits attended, study procedures undertaken, the results of those procedures, accountability records of investigational products dispensed and returned, and any adverse events experienced.

Another fundamental purpose of source documentation is to allow for reconstruction of the trial. It should therefore be a complete record and should tell the story of each participant's involvement in the trial from beginning to end. Two useful principles to bear in mind are:

'If it isn't documented, it didn't happen'

'Document what is done as well as what is not done'

Source data forms the basis for what gets transcribed into a case report form and is subsequently sent for analysis. It therefore must be retained for verification purposes to enable an independent observer to confirm the accuracy of the data on which the study outcomes are based. Source data verification (SDV) is a key quality control (QC) activity and is often performed by a monitor on behalf of the sponsor, as part of the oversight and quality management of the trial. In addition, CRFs should consider implementing procedures for QC checking of trial data. A risk-based approach is encouraged, with the level of SDV or QC checking being largely dependent on the criticality of the data. For example, 100% SDV may be required for data used to make dose escalation decisions in a first-in-human trial.

There may be times when it is appropriate to enter trial data directly into the case report form. In this instance, there is no other original record to allow for verification, and the case report form itself is considered to be the source. This should only be done if it is explicitly allowed within the approved study protocol. In addition, it is advisable to have a source data agreement or plan which outlines what constitutes the source data for all trial information.

It is important that source documents remain accessible at site at all times, as access may need to be given to a monitor, auditor or inspector. Source data should never be under the sole control of the sponsor, in order to protect against changes being made by the sponsor without the authorisation of the PI. If it is sent off site, the PI should retain an independent copy. This is reasonably straightforward for paper source documents, but can be more complicated when it comes to data submitted directly through an electronic system, e.g. from a wearable device to a central server. Such data should remain accessible by the PI, and if there is no independent copy at site, it should be under the control of an independent third party who can send a copy directly back to site without it going via the sponsor.

Regulatory Perspective

Some examples of inspection findings relating to source data are given below. Although some of these issues may appear minor, they point towards a lack of understanding of good documentation requirements. For an independent observer, such data would fail to instil confidence in data quality and patient safety. Systematic deficiencies in documentation can lead to questions about data integrity, potentially resulting in decisions to exclude data from analysis. Entire datasets may even be deemed unfit for use, wasting time and effort.

Inspection Findings with respect to Source Documentation

- Multiple records for same data points - unable to determine which served as source record, e.g. multiple versions of visual analogue scales completed for same visit with different values.
- Clinical significance for out of range lab values not documented on lab reports, or conflicting information in source documents, e.g. high glucose value marked clinically non-significant on lab report although participant was referred to primary physician for further follow-up.
- Missing pages, unexplained corrections months after initial entries, incorrect participant identifiers, incorrect dates e.g. same date for different visits.
- Numerous AEs not reported in CRFs, delays in transcribing into CRFs, discrepancies between source and CRF - lack of timely AE reporting jeopardises participant safety and data reliability.
- Incorrect / incomplete documentation regarding disposition of drugs - dates, quantity and use by participants.

In Conclusion

Source data is an essential part of study documentation and is crucial to the integrity of trial outcomes. As well as being the medical record for participants, it plays an important role in allowing verification of study data and reconstruction of the trial. To fulfil these roles it must comply with ALCOA(C) principles. Irrespective of trial outcomes, accurate documentation supports the fundamental principle of protecting the rights, safety and well-being of participants.

Ask yourself...

- *Do CRF staff understand what source data is and why it is important? Have they received any training in good documentation practices, including ALCOA(C) principles and GCP compliant corrections? Have sponsors provided training on study specific aspects?*
- *Are there written procedures that support good documentation, encompassing the recording, management, maintenance, archiving and retrieval of source documents?*
- *For trials running in the CRF, do you know what constitutes the source data, particularly if there are multiple records? Is this documented anywhere?*
- *Do PIs have appropriate oversight of source data, including regular review and resolution of issues? Are medical decisions made by the PI or a delegated sub-investigator appropriately documented in source documents? Does the PI retain control of all source data at site?*
- *Are there processes to verify the data entered into the study database? Does the sponsor or their representative have appropriate oversight of data entry, e.g. by performing SDV? Does the CRF have an internal process for quality control checking of trial data?*

4. Computerised Systems

Some general considerations for using computer systems in clinical trials are briefly outlined below. Refer to the MHRA 'GXP' Data Integrity document for more detailed guidance.

4.1 User Access and Permissions

Computer systems used to generate, process or store clinical trial data should have appropriate controls in place to ensure that only authorised personnel have access to the data, and only have permissions to carry out functions they are authorised to perform. Particular care should be taken to restrict administrator rights to the minimum number of people, and ideally people with no vested interest in the data. There should be procedures in place to remove system access from staff who no longer require it, e.g. staff who have left the organisation or no longer work on the study.

A record should be maintained of who has had access to the system and their permissions at any point in time. This may be captured by the system, but if not then a separate record should be maintained. Such records may be requested during regulatory inspection.

User access and permissions are often controlled using unique usernames and passwords, which has the additional benefit of allowing data entries to be attributable to an individual. For these reasons, the sharing of passwords or use of generic logins should be avoided whenever possible. If a system does not allow for individual logins, alternative controls should be put in place, e.g. paper records of system use.

4.2 Security and Backup

Security of trial data is important no matter what format it is in, but there are some aspects specific to electronic data. In addition to access restrictions, consideration should be given to where data is physically held and who has control of it. This is of particular importance for web or cloud based systems, or where data is held by a third party service provider. This has implications for both data integrity and data protection, and requirements including security, access and retention should be clearly documented in appropriate contracts.

A backup copy of the data, including metadata and relevant system configurations, should be retained in a separate location in case of system failure. There should be documented backup and recovery procedures in place, and these should be tested periodically.

4.3 Electronic Approvals and Signatures

From an investigator signing a protocol amendment, to a participant confirming consent on a device, it is increasingly common for signatures and approvals to be captured in an electronic format. However, there are strict legal and regulatory expectations for this, and simply inserting an image of a signature will often not be sufficient. In addition to the 'GXP' Data Integrity document, further guidance can be found in the HRA and MHRA Joint Statement on Seeking Consent by Electronic Methods, September 2018.

4.4 Computer Systems Validation

An important aspect of data integrity in the digital age is ensuring that any computerised systems used in clinical trials are fit for their intended use, i.e. they perform consistently as expected. This usually involves a series of documented checks designed to test the required functionality of the system prior to its release for use. At a basic level, this is what is meant by computer systems validation (CSV). CSV primarily applies to software developers, but users also need to be taken into account:

> *"acceptance of vendor-supplied validation data in isolation of system configuration and users intended use is not acceptable... vendor testing is likely to be limited to **functional verification** only and may not fulfil the requirements for **performance qualification**."*
>
> MHRA 'GXP' Data Integrity Guidance and Definitions, Mar 2018

In most cases, CRF staff will be system users and will not be involved in functional verification. However, if software has been installed or configured/customised by CRF staff, then some level of documented testing should be carried out to ensure it remains in a validated state. Some considerations might include:

- Has the installation been successful and is the software performing as expected?
- Is performance affected by other software installed on the same operating system?
- Have any software updates affected the integrity of data already held in the system?

On occasion, CRFs may acquire software to be used in trials directly from an eSystem vendor and will be the software owner. In this case, the CRF should satisfy themselves that the software both complies with GCP requirements and has been appropriately validated. If validation documentation can be provided, some things to check might include:

- Documentation is for the correct version of the software
- Tests covered all the functions to be used, and any test fails have been resolved
- All tests were completed before validation report sign off and product release.

If limited/no documentation is provided, and no alternatives are available, use of the system should be considered on a risk-proportionate basis. This may depend on data criticality, possible risk mitigation measures, and assurances provided by the vendor (e.g. software was developed under an accredited quality system which includes validation procedures).

Bespoke software developed in-house for clinical trial purposes will require extensive CSV, and CRFs should be aware of the expectations before undertaking this activity. The UKCRF Network QA Group has previously issued advisory notes on CSV for CRFs, with reference to further regulatory and industry guidance.

In addition, system users, as well as any supporting documentation (e.g. the user manual), are considered to be part of the overall computerised system and must also be fit for purpose. For users, that means suitable training and access/permissions have been given to allow them to use the system correctly. Documented evidence of this should be retained.

5. Audit Trails

To ensure data integrity, regulators expect all records to be generated in a timely manner and be accurate, complete, attributable, accessible and protected from loss, modification or deletion. An audit trail is key for evidencing that these expectations have been met. Auditability is the property of being able to reconstruct the actions and decisions that result in an end state in a system. An audit trail should answer the questions **who**, **what**, **when** and **why**. In effect, the Investigator Site File (ISF) is the overall audit trail for a study at a site.

Paper Audit Trails

The term 'audit trail' is strongly associated with electronic systems, but in practice the basic requirements for reconstruction of trial activities do not differ between electronic and paper records. In theory, all changes made to a paper record should be self-evident and comply with ALCOA(C) principles. The audit trail is usually demonstrated by signing or initialling and dating entries and changes. Regulators expect that paper systems will be completely auditable, with all information about changes evidenced appropriately. This will depend on a number of factors, and data integrity can only be achieved using systems that are fit for purpose, with appropriate user controls and training.

Electronic Audit Trails

Proving **who** did **what** and **when** can be achieved relatively easily with a well designed electronic audit trail, although there are some complicating factors to be considered. All data entries and modifications should be logged automatically against a user profile with a date and time. Changes should be evidenced and the old data retained. This ease of use is a primary advantage of electronic systems in terms of auditability. However, it is vital that users don't share passwords, as entries or changes will no longer be attributable.

The key element often missing from audit trails is **why**. The rationale for a change is not always recorded in the audit trail, and where this happens the change must be evidenced by an additional process. Frequently sponsors will use a data correction form (DCF) or equivalent to log the rationale for a change, but it is not uncommon for changes to be managed on a much less controlled basis and this has led to serious findings on inspection.

Another consideration is access. Reviewing the audit trail can be a useful exercise to identify otherwise hidden issues, so users should be able to view and understand it. However, many systems require an administrator to access the audit trail and interpret what it means. This can have implications for monitors, auditors and inspectors, who will often request access to the audit trail and will expect to be able to review the metadata within it.

However, care must be taken to ensure that user access to audit trails is restricted to "read only". The ability to turn off, edit or delete any part of the audit trail may cast doubt on the accuracy and attributability of the data. Any such permissions must be tightly controlled and restricted to identified authorised personnel only. Modifying the audit trail should only ever be considered in exceptional circumstances, and should be fully documented with a robust justification.

Archiving can also pose particular challenges. Systems often require proprietary software to access the audit trail, and if the system is replaced it is not always obvious how the audit trail will be evidenced. Audit trails are often databases in their own right and cannot be simply printed because the information is dynamically linked. If an electronic audit trail is required to demonstrate compliance, a key question is how it will be accessed in the future.

Regulatory Perspective

Inspection Findings with respect to Audit Trails

- Portal used for investigator data change requests had no audit trail – it was not possible to identify if changes had been made to requests after submission.
- Insufficient audit trails for Electronic Health Records and paper source documents – not possible to verify who had made entries and when, or who had made changes and why.
- Audit trail for Interactive Response Technology (IRT) system not readily accessible and did not include time stamp to verify when actions were taken.
- Insufficient audit trails in study database and paper records of data amendments.
- Data returned to investigator as PDFs after database lock, but did not include all metadata.

Findings related to audit trails are increasingly common. The implication is that regulators consider metadata such as audit trails to be equally as important as the data itself, and expect it to be maintained to a similar standard.

In Conclusion

Electronic systems are increasingly used in trials, but the basic principles of auditability are the same regardless of the format of the data. The ALCOA(C) principles that are applied to trial data are equally applicable to metadata such as audit trails. Observing these principles, and ensuring systems are appropriately validated, will help maintain data integrity.

Ask yourself...

- *Have you checked that your electronic system has an audit trail? Is it visible to you / a monitor / an inspector? If not, can it be requested from a system administrator? For older systems with no audit trail, how else could you provide this evidence?*
- *Is your system's audit trail complete – does it evidence the who, what, when and why of any entries or changes?*
- *Is your system's audit trail turned on? Can it be turned off or edited? If so, who has the user permissions required to do this?*
- *Does your electronic system retain the audit trail indefinitely? If not, how can you capture it? How will this metadata remain linked to the data it supports? If creating a printout or pdf (i.e. a "flat" file), how will this retain all the attributes of the live data (i.e. the "dynamic" file)?*

6. Electronic Health Records (eHRs)

Use of electronic health records, (eHRs), electronic medical records (eMRs) or electronic patient records (ePRs) is now commonplace. Whilst the terms are often used synonymously, eMRs/ePRs usually provide a narrower view of an individual's medical information at an organisational level. The eHR is the broader term that provides the "integrated" view of all providers involved in an individual's care. Regardless of terminology, these systems capture the clinical and other health information relating to patients that will be used as source data in relation to clinical trials, and therefore the same principles for data integrity must apply.

The Department of Health is committed to digitisation of the NHS to make it an organisation able to "operate paperless at the point of care". This strategic goal was incorporated in the Personalised Health and Care Strategy 2020. Although the ambition is a paperless NHS, hybrid systems comprising electronic and paper records are currently common. As such, it is crucial that CRFs have a clear understanding of what the source data is and where it is held.

Providers of eHRs (including in-house digital service / informatics teams who may have developed bespoke systems) may not have an understanding of the GCP requirements that must be met for systems used in the conduct and reporting of clinical trials. It is important that personnel with the relevant knowledge are consulted in defining user requirements and in the ongoing enhancement and development of the system.

The principles and standards that apply to any computer systems used in clinical trials are also applicable to eHRs. This includes physical security, restricted access, records of user access and permissions, data protection, back-up and disaster recovery, system validation and working processes for change control and system failure.

In order to ensure that regulatory and GCP standards can be met, eHRs must also be able to meet the following requirements:

- Have the ability to flag / search for clinical trial participants in the system;

- Provide read only access to clinical trial monitors, auditors and inspectors, including access to the audit trails. Access should ideally be restricted to trial participants so as to respect and protect confidentiality of patients not participating in the trial;

- Ensure that appropriate medical oversight can be demonstrated, e.g. verification of entries, review of laboratory results, etc.;

- Show when entries are made and by whom, so that the documentation provides a full audit trail of events, including any amendments or deletions and the reasons for any changes;

- Ensure that there are controls in place where data which may be in an editable format is imported from another system, e.g. radiological images. Where images require further evaluation this may include a level of modification, e.g. removing identifiable information and replacing with trial ID. The original should be retained and procedures should be in place to cover such processes;

- If data is transferred or exchanged between systems, e.g. laboratory data visualised in the eHR is generated in another source system such as a Laboratory Information Management System (LIMS), there are appropriate technical and quality controls in place to ensure the integrity of the data passed through;

- If records that have been generated on paper are incorporated into the eHR via scanning methodology, there are appropriate controls in place to ensure that the scanned image is a true copy, retaining all the attributes and information of the original. This could include a degree of quality control checking or validation of the scanning procedure;

- Ensure there is appropriate archiving procedures in place to ensure long term reliability, retrieval and reproducibility of electronic data (and metadata), in line with regulatory retention times.

Organisations who were early implementers of eHRs may not have given the appropriate level of consideration to how the translation of paper systems would be implemented and controlled, to ensure that they could effectively meet the requirements in relation to clinical trials. This may pose a risk to the integrity and acceptability of the data for clinical trial use.

Regulatory Perspective

The MHRA have previously set out regulatory expectations for eHRs in their Position Statement and Guidance - Electronic Health Records, 16 September 2015. MHRA GCP Inspectors have seen a number of issues with eHRs in relation to their compliance with clinical trials legislation and GCP principles.

Inspection Findings with respect to eHRs

- Inability to readily access audit trails - audit trails are not visible to the end user.

- Insufficient audit trails to enable reconstruction of changes to data – audit trial may indicate that a change was made but not what the change was or does not record a reason for the change.

- Insufficient audit trails – not possible to verify who had made entries and when. Lack of information relating to the reason for any changes to the data.

- Insufficient audit trails to enable verification of investigator oversight of data, such as review of laboratory results, scan reports etc.

- Lack of available documentation relating to the validation of the system.

- Insufficient consideration of quality control (QC) systems and procedures defining how the eHR system will be used to support clinical trial requirements.

- Incomplete records being printed from the eHR for the purposes of monitoring, audit or inspection.

Inspection Findings with respect to Scanned Records

- Ineffective structure and organisation of the records – e.g. scanning of paper records as PDF files in no particular order and with missing sections, thereby making trial reconstruction potentially impossible due to gaps in source data.
- Scanning is sub-contracted to companies operating to their own QC processes, without adequate checks on whether these processes are sufficient or appropriate.
- Poor resolution of scanned documents.
- Black and white scanning of colour records, resulting in the loss of associated metadata.
- Bulk scanning and subsequent disposal of paper medical records (including clinical trial source data) without having a robust process for making 'certified copies' in place. This is required to ensure that the electronic copy is an accurate copy of the source, and to enable verification of the quality of the data.

In Conclusion

Health records tell the story of an individual trial participant's involvement in the study. It is essential that data directly entered or transferred into and retained within an eHR is reliable, accurate and complete, and therefore adheres to the data integrity principles of ALCOA(C). Unreliable health related information could adversely impact both the safety of the participant and the integrity of the trial data. Observing these principles, and ensuring systems are appropriately validated, will help protect both.

Ask yourself...

- *Do you know which software applications form part of a patient's medical record within your organisation? Do you know what eHR systems are used for clinical trial purposes in the CRF?*
- *How compliant is your organisation's eHR system with GCP requirements? Have you assessed this and is this documented? How is non-compliance mitigated?*
- *Is there adequate training available to ensure CRF staff understand the relevant functionality of the system (including access to audit trails)?*
- *Are any records in the system transcribed from elsewhere, e.g. digitally dictated letters, medical annotations? How are these verified by the investigator?*
- *How will monitors / auditors / inspectors access the system? What training is provided? Can access be restricted to the clinical trial participants? If not, what mechanisms are in place to ensure complete records can be provided while protecting the confidentiality of patients not on the trial? Is this process documented?*
- *If records are printed from the eHR, what is the process for verifying that these are true copies? How and where will these records be retained?*
- *What controls are in place to ensure data integrity is maintained throughout the data life cycle? How are audit trails reviewed and by whom?*
- *Are there any eHR user groups within the organisation? Is research / the CRF represented?*

7. Electronic Patient Reported Outcomes (ePROs)

Patient report outcomes (PROs) are data reported by the patient without direct oversight by the clinical team. The classic format is the questionnaire, used to assess parameters such as quality of life, disability, symptoms, mood, etc. They may also take the form of a diary and include elements of compliance, e.g. tracking dosing and matching with IMP accountability logs. Implementing PROs using electronic systems provides advantages in terms of participant compliance and data transcription, but is not without risks.

New Platforms for Data Collection

At their most simple, ePROs are validated surveys or tools implemented using an electronic system, e.g. a tablet or smart phone. With growth in these platforms driving down cost, and wide acceptance in patient groups, sponsors are now developing apps for PRO collection. The primary appeal is improved quality of data collection and reduction in the complexity of managing paper records and transcribing into a central database. ePRO platforms not only improve compliance, but can also capture an electronic audit trail. In terms of data integrity, this offers an advantage as it allows assessment of timeliness and attributability.

Perhaps the most common way of managing ePROs is a leased tablet or smart phone provided by the sponsor, which exclusively provides access to the relevant systems for data collection. This allows the sponsor to control validation of the system and the device. Alternative models include “bring your own device”, which is attractive as it removes the need to manage leasing, but it presents significant challenges in terms of validation.

Regulatory Perspective

While specific regulatory guidance is limited, the data integrity challenges presented by ePROs are similar to other ‘e’ systems. The MHRA have provided inspectors’ perspectives on ePROs and often include these systems within the scope of ‘eSource’. Critical findings involving ePROs have been reported, e.g. implementation of a system in such a way that data modifications could not be verified, and a failure of user acceptance testing combined with incomplete audit trails resulting in a large number of records becoming unreliable.

Inspection Findings with respect to ePROs

- Insufficient eDiary audit trails – not possible to verify who had made entries and when.
- Incorrect data from participant eDiaries analysed as data could not be changed in system, even though changes had been requested by investigators.
- Data submitted by participants through eDiaries, including key study endpoints, had been changed, but there was insufficient source data to support the changes.
- Lack of validation documentation for user acceptance testing of eDiary – no evidence that plan was approved, that steps were followed as per plan, or who had carried out the testing.

Site-Specific Challenges

For trial sites, many of the challenges presented by ePROs will be largely practical. It may be necessary to store a device in-clinic for the duration of a trial, as well as ensure a power supply and network connection. The workload represented by responding to data queries from the electronic system may also be a consideration.

One data integrity risk is staff training participants on the use of a system. The quality of data collection may therefore depend on the quality of training. Further risks may arise if staff are required to interact with the data, e.g. by downloading or transferring data from a device to a central database. Sites should be satisfied that the method of transfer has been appropriately validated, and should check to confirm ePROs comply with data protection requirements and local information governance policies, particularly when international sponsors are collecting data. In addition, source data, including data from an ePRO device, should remain accessible at site, and should never be under the sole control of the sponsor.

Sponsoring a Trial Involving an ePRO System

The main data integrity risks of ePROs are usually born by the sponsor, primarily validation of the system. Critical findings relating to ePROs have resulted from compound failures of validation processes. In practice, it is common for sponsors to contract a third party to develop an ePRO. These activities may be delegated, however a process for qualifying the vendor, as well as evidencing validation of the system, will be required. Commissioning of these systems can therefore be highly resource intensive and require a long period of time.

In Conclusion

The use of ePROs is a growth area in clinical trials. ePROs offer advantages for data integrity by reducing transcription errors, improving compliance and providing audit trails. However, inspections have shown they can also present a significant risk for sponsors if the system is not appropriately validated. Sites may see increased resource requirements for training staff and participants on new systems, and also need to consider infrastructure requirements.

Ask yourself...

- *Is the CRF involved in trials that use an ePRO? What is the level of CRF staff interaction with the ePRO? Are you satisfied that any associated software has been appropriately validated?*
- *Do CRF staff train participants in the use of ePROs? Is there evidence that they have received training to do this? Are you confident in the training given?*
- *If data from an ePRO is sent directly to a central server, what access does the PI / site have to this source data? Is it ever under the sole control of the sponsor?*
- *If you are the sponsor, have you ensured that ePROs have been appropriately validated prior to use? Do you have evidence of this?*
- *If you have contacted a third party, do you have evidence of vendor assessment? Does the vendor understand GCP requirements? Have you been provided with evidence of validation?*

8. References and Resources

The resources listed below have either informed these advisory notes, or are additional useful sources of guidance on aspects of data integrity in a GXP environment.

General Guidance

- GXP Data Integrity Guidance and Definitions, MHRA, Revision 1, Mar 2018
- ICH E6(R2), Guideline for Good Clinical Practice, Nov 2016
- The Medicines for Human Use (Clinical Trials) Regulations, SI 2004/1031, as amended
- World Health Organisation Technical Report Series, No. 996, Annex 5: Guidance on Good Data and Record Management Practices, 2016
- Good Clinical Practice Guide, MHRA, 2012
  In particular, see Section 8: Data Management and Section 11.5: Study Documentation
- MHRA Inspectorate Blog. In particular, see:
  - MHRA GXP Data Integrity Guidance: Part 1 - A GCP Perspective, Mar 2019
  - MHRA's GXP data integrity guide published, Mar 2018
  - Too much pressure: a behavioural approach to Data Integrity (Parts 1&2), Mar 2017
- MHRA GCP Inspection Metrics Reports

General Guidance on Electronic Systems

- Reflection Paper on Expectations for Electronic Source Data and Data Transcribed to Electronic Data Collection Tools in Clinical Trials, EMA/INS/GCP/454280/2010, Jun 2010
- EudraLex Volume 4 Good Manufacturing Practice, Annex 11: Computerised Systems

Guidance on Electronic Consent

- Joint Statement on Seeking Consent by Electronic Methods, HRA & MHRA, Sep 2018
- MHRA Inspectorate Blog: eConsent, Oct 2018

Guidance on Computer Systems Validation

- Computer Systems Validation Advisory Notes, UKCRFN QA Theme Group, v1.0, Jul 2016
- Good Clinical Practice Guide, MHRA, 2012, Section 14.5: Computer System Validation
- MHRA Inspectorate Blog: Computer System Validation - GCP, Apr 2017

Guidance on Electronic Health Records

- Position Statement and Guidance, Electronic Health Records, MHRA, 16 Sep 2015
- Personalised Health and Care Strategy 2020, Department of Health, Nov 2014
- Good Clinical Practice Guide, MHRA, 2012, Section 11.5.2: Electronic Health Records
- MHRA Inspectorate Blog: Electronic Health Records, Jul 2019

Guidance on Electronic Patient Reported Outcomes

- MHRA Inspectorate Blog: ePRO – An Inspector's Perspective, Jul 2016

UKCRF Network QA Theme Group Members

These advisory notes were developed by the following members of the UKCRF Network QA Theme Group:

- Kirsty Adams, QA Manager, Leonard Wolfson Experimental Neurology Centre, UCL
- Kathryn Betts, QA Manager, Surrey CRC
- Jacob Bonner, QA and Governance Manager, NIHR Imperial CRF
- Jacqueline Bramley, QA Lead, Lancashire CRF
- Georgia Bullock, QA Manager, NIHR/Wellcome Trust King’s CRF
- Anna Chapman, QA Manager, Cambridge CRF
- Lucy Cooper, QA Manager, Alder Hey CRF
- Fiona Cregg, Quality & Regulatory Affairs Manager, HRB CRCI
- James Gibson, QA Lead, Edinburgh CRF (*Chair*)
- Terese Hale, CRF Nurse Manager (QA), Cardiff CRF (*Deputy Chair*)
- Lynn Hope, Quality Lead Research, The Christie Research
- Greg Langton, QA Manager, Leeds CRF
- Kim Lee, Acting QA Lead, NIHR/Wellcome Trust Southampton CRF
- Jo O'Neill, QA Manager, NIHR/Wellcome Trust Birmingham CRF
- Alex Powell, QA Lead, NIHR Guy's and St Thomas' CRF
- Carole Schilling, Quality & Regulatory Affairs Manager, RCSI CRC
- Katy Shortland, QA Lead, Sheffield CRF
- Eilidh Wright, QA Lead, Glasgow CRF

Acknowledgements

The UKCRF Network QA Theme Group would like to acknowledge the contribution and assistance of the UKCRF Network Strategic Management Team.