UKCRC Registered Clinical Trials Units

UKCRC Registered CTU Network – Data Cleaning and Query Management

UK CLINICAL RESEARCH COLLABORATION

igniting our potential

Data Cleaning and Query Management

Editor

Amanda Loban – University of Sheffield Clinical Trials Research Unit

Authors:

Paigan Aspinall – Peninsular Clinical Trials Unit

Chloe Bruce – Brighton and Sussex Clinical Trials Unit

Gayle Dower – Royal Marsden Clinical Trials Unit

Jacqui Gourlay – Glasgow Clinical Trials Unit

Patrick Julier – Oxford Clinical Trials Research Unit

Amanda Loban – University of Sheffield Clinical Trials Research Unit

David Murray – National Perinatal Epidemiology Unit

Kate Sparksman – PRIMENT Clinical Trials Unit

Jo Steele – Papworth Clinical Trials Unit

Sarah Sumpter – Hull Health Trials Unit

Stephen Townsend – MRC Clinical Trials Unit at UCL

Val Wadsworth – York Trials Unit

Ruth Wood – Newcastle Clinical Trials Unit

Contents

1. Introduction

The purpose of this document is to provide guidance on cleaning of data collected within a clinical database for clinical research. The document has been developed with consideration to regulatory requirements and CTU high level processes. This document is designed to be used as recommendation for good practice and is not exhaustive. It is intended to be used as a starting point for new CTUs and / or alongside your unit specific guidance and SOPs.

Different trial units will have different approaches depending on the tools they are using and the type of data included in the trials. The processes may be manual, automated, or a combination / hybrid approach.

This document incorporates recommendations based on the experience of the authors, representing thirteen UKCRC registered clinical trials units.

2. Summary

The approach to data cleaning will depend upon several factors including the CDMS used, the CTU SOPs and the risk assessment. Data cleaning is a key activity for ensuring data quality and integrity, including the detection of potential safety and protocol compliance issues. When carried out regularly and thoroughly it can enable the early detection of systematic issues and therefore, as well as identifying data to be corrected, it can help identify process improvements for the trial overall; with the potential to improve recruitment, protocol adherence, safety and overall performance of sites. It is therefore essential that a systematic approach to data cleaning is taken, and it is fully documented and monitored. All documentation related to data cleaning forms part of the essential documentation for the trial[1].

The process is likely to be led by members of the data management team, though this will depend upon the CTU; it will require cross disciplinary involvement potentially including data programmers, statisticians, trial managers, investigators and others as required.

A comprehensive plan for data cleaning is essential, and this guidance aims to inform the contents of a CTU specific plan. Although the aim is to make this guidance useful and comprehensive from a CTU perspective, further reading is required, section 8.5 Data validation of the MHRA Good Clinical Practice Guide[2] is recommended reading.

3. Scope

This document is focused on checking individual data items for potential discrepant data, e.g. missing, inconsistent and implausible data. It also includes checking data for protocol non-compliances, e.g. eligibility criteria, drug dosing, SAE reporting timelines, and visit windows. Although the intent of this document is not to overlap with central monitoring activities, reporting on data cleaning activities is considered in scope; this reporting is useful for central monitoring.

[1] Regulation 31A (4) of SI 2004/1031

[2] MHRA Good Clinical Practice Guide. Twelfth impression 2021

Data cleaning is an activity usually led by members of the data management team and / or by processes implemented by data management or data programmers.

3.1. In Scope

The activities covered include:

- Point of entry checks (PoE): automated data checks triggered at the point of entry. This includes the use of appropriate data types and database design.

- Post-entry checks: checks that occur after data entry has taken place. These can be performed in the Clinical Data Management System (CDMS), on exports of participant-level data, or a combination of both. Post-entry checks may be automated and / or manual and are usually performed by statisticians, data managers or programmers.

These activities lead to the flagging of potential issues (i.e. data discrepancies) and queries which require resolution.

3.2. Out of Scope

Double data entry (DDE) / Data entry verification (DEV). In general, this is an almost obsolete process with a move to EDC and away from paper, therefore paper only data collection is out of scope. Though there are some units or studies where this is still appropriate it is outside the scope of this guidance.

Source data verification (SDV). This is outside the scope and requires access to source data at site. At many units the data management plan and data monitoring plan are separate, and monitoring is not always within the data management remit.

Central monitoring, which involves the evaluation of accumulating data (or lack thereof)[3], should be considered distinct from data cleaning and is not within the scope.

Preventing the editing of data once cleaning activities are considered complete is not included within the scope; if functionality exists within the CDMS to prevent editing of data or otherwise flagging data as 'complete' (often termed 'freeze' or 'lock') individual CTUs should consider how to use this functionality. It is important to have a 'lock' process or mechanism for detecting that data has changed post 'complete'. This process should be documented, ideally within standard operating procedures (SOPs). The lock process will be covered in more detail in other guidance documents.

4. Where does this document sit within the study lifecycle?

Guidance is being developed which describes the study lifecycle. Once this is published a link to this document will be added here.

[3] Making a distinction between data cleaning and central monitoring in clinical trials - Sharon B Love, Victoria Yorke-Edwards, Carlos Diaz-Montana, Macey L Murray, Lindsey Masters, Michelle Gabriel, Nicola Joffe, Matthew R Sydes, 2021 (sagepub.com)

5. Definitions

What we describe here is for the purpose of this document only, different units may use different terminology and different combinations of checks. The purpose of this section is to describe the scenarios we are discussing.

Checks - the question, rule or check implemented to validate the data, e.g. data must be present, data must be within the predefined range (see 'Types of checks' section).

PoE (point of entry) - checks that are run in 'real-time', as data is being entered, which identify discrepant data for the attention of the person carrying out the data entry. They may be 'hard' or 'soft'

- **Hard** - point of entry checks which prevent erroneous data from being entered; to be restricted to impossible data (e.g. biologically impossible values) and data formats. Data can't be entered if it triggers a check. NB care should be taken to ensure these are wide enough to allow edge cases and prevent the risk of sites fabricating data to fit.

- **Soft** - point of entry checks which do not prevent entry of 'suspect data' but trigger a warning in real-time at the point the data is being entered. These may be actioned immediately: discrepant data may be amended (which would mean the trigger for the warning is then removed), OR data unchanged but the warning itself responded to (e.g. a reason for missing or values outside expect range, etc given), OR the warning is retained for resolution at a later time.

Post Entry - checks that are run to check for 'suspect data' after the data has been entered, i.e. run on data already saved in the CDMS. This may also refer to some checks run at the point of entry but where staff are unable to resolve immediately. These may be in-built in the CDMS, programmed outside of the CDMS or require manual review.

- **In-built** - checks which are built within the CDMS and run in batch (i.e. using a tool which runs within the CDMS); these would usually be run automatically.

- **Externally programmed** - checks which are programmed outside the CDMS, e.g. in SPSS, STATA or SAS. Ideally these would be run automatically, i.e. scheduled.

- **Manual** - checks which cannot be fully programmed, often as they require review of entered free text data, e.g. adverse events vs concomitant medications; though it may also be because of limitations with the tools used to program checks. This would include checks carried out via reports and other types of data review, such as medical review. This may also refer to a manual review of the data for discrepancies which may have been missed by the programmed checks (though if found these would ideally be added as rules to be included with the programmed checks).

5.1. Types of checks

- Data type (date, numeric, codelist).

- Impossible ranges (numeric / dates).

- Mandatory fields incomplete.
- Branching logic / dependencies - form level and field level.
- Missing fields (non-mandatory).
- Missing forms.
- Out of range (possible but unlikely values; e.g. lab values, demographics, times).
- Out-of-range visit dates (to check for systematic issues or protocol non-compliance).
- Incongruent values (e.g. two fields collecting the same data disagree (e.g. date of death in two places), best to avoid this if possible (e.g. by deriving a field, such as carrying height across from baseline to all subsequent follow ups for calculating BMI); a question answered no but information provided, best prevented by using branching logic so information cannot be entered if question is answered no).
- Longitudinal data changes (i.e. large or unexpected changes in weight over time; disease assessment / RECIST over time).
- Cross-checks with eligibility criteria.

6. Regulatory requirements

ICH GCP is a recognised international standard for the design, conduct, safety and reporting of clinical trials. The principles of the ICH GCP are adopted by the MHRA which is the UK governing body for the conduct of clinical trials. The ALCOA principles referred to in the Regulations[4] define best practice for data management in that data should be attributable, legible, contemporaneous, original record, accurate, complete, consistent, enduring and available.

"The guidance refers to ALCOA rather than ALCOA+". "ALCOA was historically regarded as defining the attributes of data quality suitable for regulatory purposes. The '+' has been subsequently added to emphasise the requirements. There is no difference in expectations regardless of which acronym is used since data governance measures should ensure that data is complete, consistent, enduring and available throughout the data lifecycle".

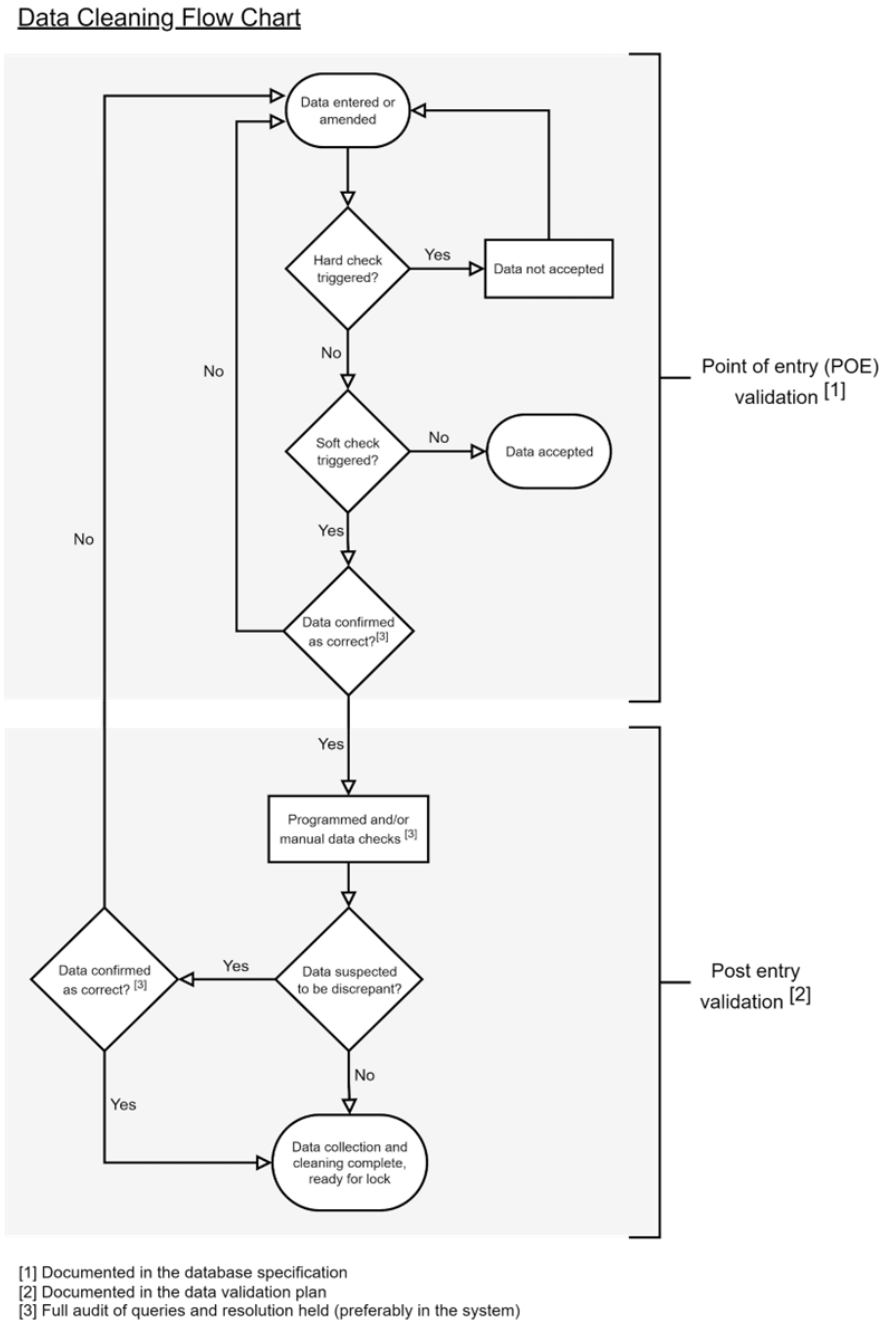
"A procedure should describe the actions taken if data review identifies an error or omission. This procedure should enable data corrections or clarifications to provide visibility of the original record, and traceability of the correction, using ALCOA principles".

[4] MHRA GXP Data Integrity Guidance and Definitions; Revision 1: March 2018

7. Process

7.1. Flow chart

The following flow chart gives an overview of the data cleaning in scope.



[1] Documented in the database specification
[2] Documented in the data validation plan
[3] Full audit of queries and resolution held (preferably in the system)

7.2. eCRF / database design

The data manager should be involved in the early stage of CRF development and database design (see also Guidance on the Development and Usage of eCRFs / CRFs).

The use of appropriate formats (e.g. numeric fields for numbers, date fields for dates, drop-downs lists etc), ensuring data is collected at the right visit, preventing duplication of data, liaising with stakeholders to confirm correct lab ranges, etc are all important early stages to the design.

7.3. Defining and documenting checks

It is recommended that all checks (and the methodology for implementing) are documented and approved by relevant personnel (depending on the unit this may include members of the data management team, data programmers, statisticians, trial managers, investigators and others). The document used for this will differ between units and will be detailed within SOPs, examples are Data Validation Specification (DVS), Data Validation Plan (DVP), Data Collection Specification, Data Management Plan (DMP), Database Specification, Data Dictionary. Point-of-entry checks are incorporated within the build and therefore most likely documented within the database specification documents. Henceforth this documentation will be encompassed by the generic term ‘data validation specification’.

The checks should be relevant to the protocol; the documentation should be version controlled and updated in line with any protocol changes.

Other considerations include the data source, category of data and the trial risk assessment.

7.4. Data source

Trials typically consist of data from several sources, i.e. that:

- collected by site staff (site data) typically during a contact with the participant; the source may be the participant notes or the CDMS (if EDC and not documented elsewhere).
- collected directly from participants via electronic or paper surveys (questionnaire data), typically Participant Reported Outcome Measures (PROMS).
- imported from machines (electronic data), e.g. MRI data, accelerometer data etc.
- from routinely collected data sources (routine data), e.g. HES data. Refer also to ICH E6(R3) Guideline – Annex 2 which specifically addresses Good Clinical Practice (GCP) considerations for clinical trials incorporating Real-World Data (RWD).

The approach to data cleaning will likely differ depending on the data type.

It is likely that PROMS will require much less cleaning, as if entered directly by participants there may not be recourse to follow-up. However, reporting on completeness could help to identify systematic errors (e.g. missing pages, lack of clarity, data entry issues (if paper completion followed by central data entry): these may be better addressed in missing data reports, rather than via data cleaning.

Electronic data and routine data sources are likely to be subject to minimal checking, to identify missing information and potential systematic issues with the machines / data collection method.

Site data is likely to be subject to most scrutiny, this cleaning can also help identify potential misunderstandings, protocol non compliances, resourcing issues (i.e. data not being entered into the CDMS in a timely manner) and data inconsistency.

7.5. Category of data

Categories may include

- 'Critical'[5] e.g. data that is important for eligibility, the primary endpoint, secondary endpoints, and safety.
- other endpoint data.
- data evidencing protocol and GCP compliance.
- data to support the management of the study rather than the analysis e.g. telephone contract log.

Any decisions around the approach to cleaning based on data source and category should be informed by the trial risk assessment, documented and approved.

7.6. Approval of checks

As discussed, the checks to be included (and excluded) should be documented and approved by appropriate members of the study team, as defined with unit SOPs.

7.7. Programming / Creation of checks

These may be in-built in the system or may be built outside the system e.g. Statistical or database software.

7.8. Testing and release of checks

The edit checks configured from the 'data validation specification' must be tested prior to implementation for use on live data. This testing must be documented, as should any changes to these checks (see change control).

There are different potential approaches to testing and documentation. The approach will depend on several factors including the systems used; and the trial risk assessment informed by the available resources. The approach should be documented in SOPs.

[5] ICH GCP E6 R3 refers to these as critical to quality factors

7.9. In-built checks

These checks are often built during the set-up of the database, often these are point of entry checks, but this may include checks which are run post entry. These checks may be built using the 'native functionality' of the CDMS or using in-built or add-on tools / applications; they appear in-built for the end user.

Typically, all checks must be defined, documented and built prior to starting testing. Checks should be tested before the build is approved as part of the database testing process.

Recommendations for types of testing and documentation include:

- The tester entering test data that should trigger the check (boundary testing for upper and lower limits and considering all parameters). This testing must be documented. Documentation may consist of:
    - Annotating a comprehensive list of requirements in a test plan (in conjunction with test data): i.e. documenting the outcome: pass / fail, including evidence (in accordance with the documented approach).
    - More generic sign off to confirm all general and specific tests in the test plan / checklist have been carried out and work as expected, this would need to be in conjunction with guidance regarding how to carry out these tests.
- Reviewing system source code vs the database specification to ensure all tests are incorporated and correct.
    - Can be useful but not a common approach, it would be carried out on a risk assessed basis and by staff appropriately qualified and trained.
    - Documentation is likely to consist of the annotated source code.
- Running automated test scripts / unit testing.
    - The output from the automated test scripts provides the documentation

In all cases an approval form confirming the testing has been completed successfully and the database is ready for release is recommended.

Ideally, these tests should be carried out by an independent tester, i.e. someone other than the person(s) who built the checks in the CDMS.

7.10. Externally programmed

These checks may be performed by statisticians, data managers or programmers. Consider how to validate these checks and how to document this, e.g.

- Double programming, ideally using different software (e.g. SAS vs R).
- Code review.
- Manual review of output vs test data.

Consider how often and when this validation would be carried out

- Per generic / reusable piece of code (e.g. macros for each check type).
- Per study, when the checks are initially programmed (and revised).
- Every time the checks are run, i.e. each time output is generated (not recommended).

Consider also taking (and documenting) a risk-based approach, based on the trial risk assessment, to justify the validation method adopted.

Maintaining a repository of standard forms, fields, checks, and scoring algorithms; for which documented testing has already been performed could lessen the burden of testing and documentation by re-using these elements with a reduced need for testing. The level of testing should be appropriately justified and requires careful consideration as part of the trial risk assessment.

7.11. Change control

Any necessary changes to checks, after they have been released, should be introduced in a controlled way and the document recording the checks kept up to date and version controlled; this includes changes to checks programmed outside the system.

Necessary changes requiring change control include, for example: correcting checks that are misfiring (i.e. where valid data is being flagged as discrepant), addition of new checks identified following a data review or removing unnecessary checks where the study team agree.

Consider how requests for changes are reviewed, how they are assessed for risk, potential impact on existing data, how they will be applied to existing data (i.e. new data only or all existing data and whether existing queries that are no longer triggered are closed or deleted), and how this review is documented. If changes are agreed, document the implementation, testing, and release in accordance with CTU SOPs.

Depending upon the nature and risk of the change, elements of the processes outlined above will be followed.

7.12. Distribution and tracking of queries

Documenting the data cleaning activities, as well as accompanying changes to the original data, is critically important for maintaining data integrity and quality, including demonstrating adherence to ALCOA+ principles.

Data cleaning activities will generate queries which should be tracked to appropriate resolution, providing evidence that data cleaning has been applied consistently and successfully, and ensuring the process is clearly documented.

Tracking of queries is important to demonstrate data cleaning has been carried out appropriately. Ideally this tracking would include information regarding when and how the query was generated (e.g. manually or through the CDMS), who raised the query, the status

of the query (e.g. open, resolved, closed), and if an explanation is required to close the query, the final explanation, who closed it and when it was closed.

Tracking of queries is likely to be most effectively and efficiently managed within the CDMS system. It is therefore recommended that data discrepancies identified outside of the system are manually raised as a query within the CDMS. Sites should also respond to queries through the CDMS.

It is often judicious to export lists of queries and send as reports to the sites to support them in engaging with the data cleaning process. Sites should be encouraged to respond via the CDMS (assuming the tracking information is embedded as outlined above). On occasion sites may provide explanations to close the queries by email or telephone; especially where time is limited. In these circumstances it may be reasonable for CTU staff to close the query regarding discrepant data and include the explanation within the tracking system (in accordance with CTU SOPs). It is recommended that relevant correspondence is referenced within the explanation in the tracking system; with correspondence filed in accordance with the CTU SOPs.

It is recommended that a record of data cleaning activities is maintained which evidences the data cleaning that has been performed, including those which have not generated queries. The CDMS query tracking system may provide this functionality, or additional documentation may be required, especially if there are ad hoc data cleaning activities that are not otherwise recorded.

There are two main resolutions for a query: (1) the data is updated and no longer discrepant or (2) an acceptable explanation is given for the discrepant data.

Any updates to data must be documented in the audit trail for the data: the audit trail must store what was changed, by who, when and if required why.

8. Self-Evident Corrections (SECs)

For trials that are entered by staff at investigator sites, SECs are generally considered unnecessary.

A 2021 MHRA blog[6] stated: "There should be no need for any pre-authorised 'self-evident corrections' to be made by sponsor data management as occurs for paper CRFs and a sponsor database because any changes to an eCRF should be routed back to the investigator as a query and the paper-based system did not involve the sponsor staff amending the CRF itself, just their database copy of it. It is not necessary for the sponsor staff to have edit access to the eCRF forms."

It is not recommended that sponsors / CTU staff have edit rights to change the eCRF as any changes required should be directed to the investigator as a query. The same MHRA blog6 as quoted above also states "The investigator should authorise the access of any person assigned editing rights in the eCRF and maintain oversight of this where the sponsor controls it, as well as authorisation *of any entries and changes to the data by having visibility and oversight of such changes (ICH GCP 4.9.3)".*

[6] Is your eSystem actually an eCRF (electronic case report form)? - MHRA Inspectorate (blog.gov.uk)

Though paper CRFs are considered out of the scope of this guidance, for clarity reference to the guidance in the MHRA 'Grey Book'[7] (Section 8.5.6 Self-evident corrections) is provided here as it is of relevance:

"Sometimes the sponsor may have a 'self-evident correction' or 'assumptions' document. This lists changes that can be made to the data by data management personnel without specific referral to the investigator with a data query. An example would be the correction of the common error that occurs in early January when dates are written using the previous year. Such an agreement can reduce the effort to correct obvious mistakes and therefore the time that both data management and the investigators spend on data corrections. This document should be approved by the investigator prior to the data management activities taking place. Any changes to the document should be approved by the investigator prior to implementation.

At the end of the trial, the investigator should be provided with details of the actual changes made using the process, which is often not done. It would therefore be clear in the database which data edits had been made as a result of self-evident corrections. The use of self-evident corrections is expected to be limited and is certainly not a substitute for a formal data query process, particularly when under time pressure."

9. Other considerations

9.1. Data cleaning reports and monitoring

Monitoring the data cleaning process, including reviewing both the quantity and types of queries raised, provides important information regarding data quality and integrity. This information increases the chance of early detection of systematic issues and the likelihood that effective mitigation can be taken. It is therefore important to consider reporting activities around data cleaning, for central monitoring and for reporting to the sites and oversight committees.

For oversight committees it is useful to provide information regarding the total number of queries ever raised, displayed as open and closed queries, generally on a by site basis as a proportion of data entered (ideally with an indication of the length of time they have been open), along with details of missing data if not already captured within the open query information. This information should be used to consider if there are any potential systematic issues that may need further action, e.g. additional resources or training.

Detailed guidance on data reporting is outside the scope of this guidance and will be covered within a separate guidance document.

Plans for follow-up of data queries with sites should be documented, most likely within the DMP, DVP or monitoring plan, including who follows these up, how they are followed up and how frequently to:

- run checks (especially where ad hoc).
- review open queries.
- follow up with sites.

[7] MHRA Good Clinical Practice Guide. Twelfth impression 2021

9.2. Standards

The use of standard forms has been introduced in the 'Testing and release of checks' section. Other guidance documents will also cover the usefulness of using standards as they save time at all stages: CRF design; database build; database validation; defining and testing validation rules; data reporting; and analysis. These standards would ideally include data formats, codelists, inbuilt checks, branching logic, coding, scoring algorithms and other applicable checks (such as standard ranges for standard items (e.g. adult height and weight)). CTUs may develop their own data standards and / or use industry standards (e.g. those put forward by the Clinical Data Interchange Standards Consortium (CDISC)).

9.3. Medical coding

Checks to aid medical coding (e.g. adverse events and concomitant medications) may be carried out as part of data cleaning and included as part of the processes described above. These checks help ensure the information given can be coded accurately, e.g. ensure only single events are reported per form; check symptoms are not coded instead of a diagnosis, where a diagnosis is available (e.g. if the AE/SAE of bowel obstruction is recorded the symptoms of pain, constipation etc should not be coded). It is important to record, most likely within the DMP, the dictionary(ies) and version used (e.g. MedDRA, CTCAE), as well as the process followed. GCP indicates that medical coding is 'undertaken or at least reviewed by a medically trained professional'. It is also recommended that this review is documented.

In some CTUs this task might fall to other members of the team (e.g. trial managers) and / or may be recorded externally from the CDMS, it may be out of scope of data cleaning.

Appendix 1 – Acronyms & Glossary

Term/acronym	Full Name	Description	References
ADaM	Analysis Data Model	Standards for organising and submitting clinical trial data to regulatory authorities	CDISC | Clear Data. Clear Impact.
ALCOA	Attributable, Legible, Contemporaneous, Original, Accurate	Defines the attributes of data quality suitable for regulatory purposes	MHRA GXP Data Integrity Guidance and Definitions; Revision 1: March 2018
CDASH	Clinical Data Acquisition Standards Harmonisation	Standards for the collection of clinical trial data	CDISC | Clear Data. Clear Impact.
CDISC	Clinical Data Interchange Standards Consortium	Widely accepted and used data standards in clinical trials	CDISC | Clear Data. Clear Impact.
CDMS	Clinical Data Management System	A specific clinical database to manage the data of a clinical trial	
CTCAE	Common Terminology Criteria for Adverse Events	A tool for reporting and grading adverse events in cancer clinical trials.	Common Terminology Criteria for Adverse Events (CTCAE) | Protocol Development | CTEP (cancer.gov)
DDE	Double Data Entry	A process to ensure data is accurately entered by having two staff enter the data separately, to enable comparison using software to highlight differences.	
DEV	Data Entry Verification	A process to ensure data is accurately entered by verifying it matches the paper form.	

Term/acronym	Full Name	Description	References
DMP	Data Management Plan	An overview of the data management process to be applied to the trial to assure data quality.	MHRA Good Clinical Practice Guide Twelfth impression 2021
DVP	Data Validation Plan	A plan for how the data will be validated (maybe be part of the DMP). The plan may cover roles and responsibilities, the types of checks, how they will be chosen and documented, the processes used for implementing the checks and how any problems will be resolved.	MHRA Good Clinical Practice Guide Twelfth impression 2021
DVS	Data Validation Specification	The description of the checks to be performed.	MHRA Good Clinical Practice Guide Twelfth impression 2021
EDC	Electronic Data Capture	Entering data into an electronic case report form (eCRF)	
GCP	Good Clinical Practise	A set of internationally recognised ethical and scientific quality requirements that must be followed when designing, conducting, recording and reporting clinical trials that involve people.	Good Clinical Practice - Health Research Authority (hra.nhs.uk)
HES	Hospital Episode Statistics	A curated data product containing details about admissions, outpatient appointments and historical accident and emergency attendances at NHS hospitals in England.	Hospital Episode Statistics (HES) - NHS England Digital
ICH	International Council for Harmonisation of Technical Requirements for	An organisation bringing together the regulatory authorities and pharmaceutical industry to discuss scientific and technical aspects of pharmaceuticals and develop ICH guidelines.	ICH Official web site: ICH

Term/acronym	Full Name	Description	References
	Pharmaceuticals for Human Use		
MedDRA	Medical Dictionary for Regulatory Activities	A rich and highly specific standardised medical terminology to facilitate sharing of regulatory information internationally for medical products used by humans.	English | MedDRA
MHRA	Medicines and Healthcare products Regulatory Agency	The regulator of medicines, medical devices and blood components for transfusion in the UK.	About us - Medicines and Healthcare products Regulatory Agency - GOV.UK (www.gov.uk)
PoE	Point of Entry	Fully described in the 'Definitions' section	
PROMS	Participant Reported Outcome Measures	Usually, a questionnaire completed by the participant to assess their health status (e.g. quality of life, symptoms etc)	
RECIST	Response evaluation criteria in solid tumours	A set of published rules that define when tumours in cancer patients improve ("respond"), stay the same ("stabilise"), or worsen ("progress") during treatment.	RECIST (eortc.org)
SAE	Serious Adverse Events	'Any untoward medical occurrence or effect that at any dose results in death, is life-threatening, requires hospitalisation or prolongation of existing hospitalisation, results in persistent or significant disability or incapacity, or is a congenital anomaly or birth defect'	Article 2(m) of Directive 2001/20/EC
SAS	Statistical Analysis Software		Leading Statistical Analysis Software, SAS/STAT | SAS

Term/acronym	Full Name	Description	References
SDTM	Study Data Tabulation Model	A standard format for organising and submitting clinical trial data to regulatory authorities.	CDISC | Clear Data. Clear Impact.
SDV	Source Data Verification	A process employed in clinical trials to check data against source documents for accuracy and completeness.	
SECs	Self-Evident Corrections	Changes made to discrepant clinical data without sending a query to the site in order to correct the data based on incontrovertible supporting information entered on the Case Record Form (CRF).	
SOPs	Standard Operating Procedures	Instructions that describe how to carry out a specific process in accordance with the relevant regulations.	
SPSS	Statistical Product and Service Solutions	A widely used statistical software for data analysis in social science, business, and healthcare research	
UKCRC	UK Clinical Research Collaboration	The UKCRC Registered CTU Network is an independent organisation dedicated to providing support for UKCRC Registered CTUs undertaking non-commercial and investigator-led clinical trials both in the UK and overseas.	UKCRC (ukcrc-ctu.org.uk)