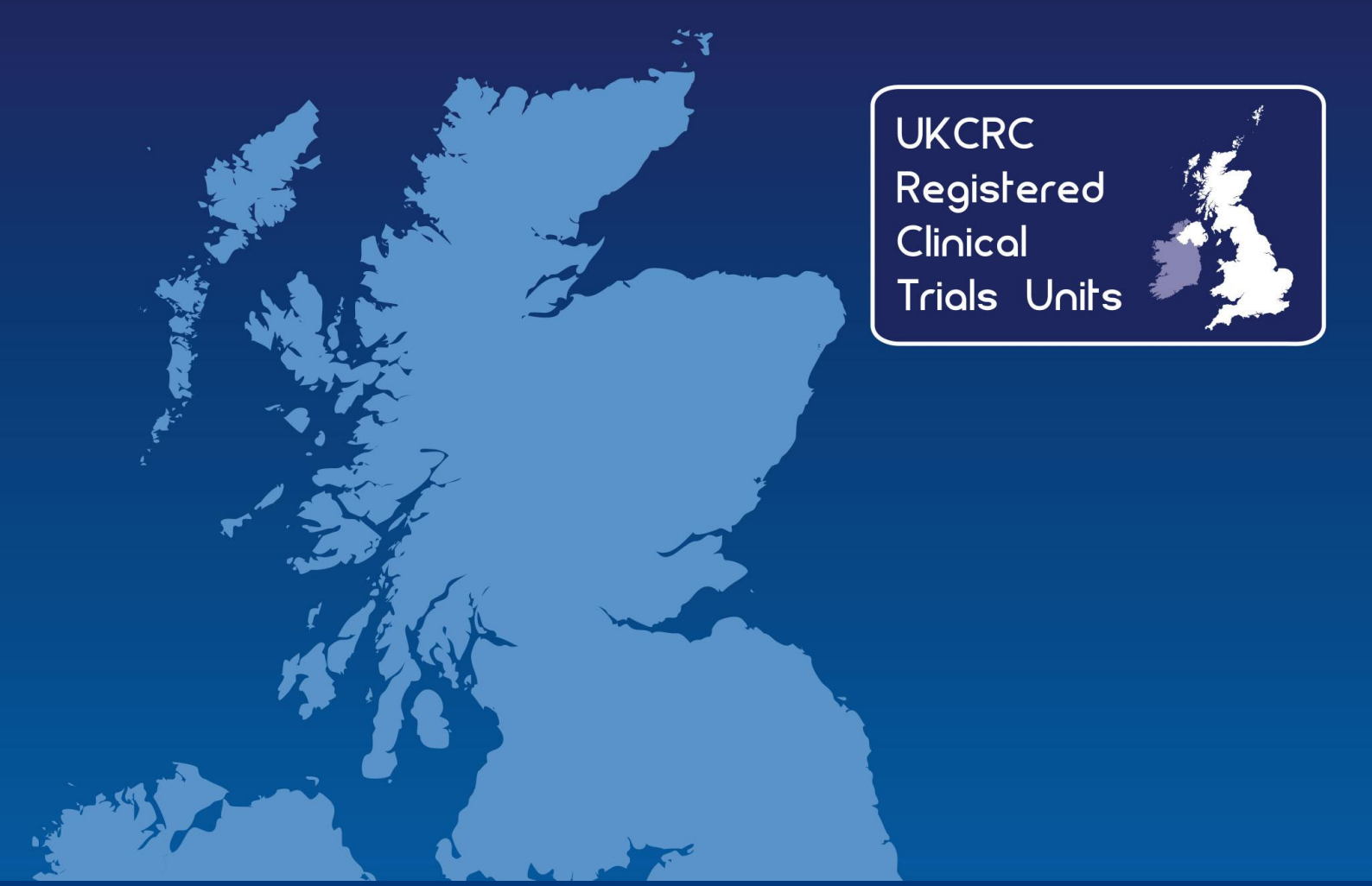

UKCRC Registered CTU Network – Data Reporting Guidance

Data Reporting Guidance

Authors:

Tim Chater – Liverpool Clinical Trials Centre

Amanda Loban – University of Sheffield Clinical Trials Research Unit

Lindsey Masters – MRC Clinical Trials Unit at UCL

Acknowledgements:

This document was reviewed by the UKCRC Registered CTU Network's Chairs' Group (i.e. the Chairs of each operations group within the Network), and members of the Data and IS Operations Group.

Contents

1. Introduction

Data reports can be useful throughout the duration of a study to provide information to oversight committees such as Trial Management Groups (TMG), Trial Steering Groups (TSC), Data Monitoring Committees (DMC)[1], as well as for facilitating central monitoring, data cleaning, and safety reporting.

The UKCRC Registered CTU Network's (Network) Data and Information Systems Operations Group (DISOG) identified a lack of best practice guidance for ongoing data reporting in clinical trials. This document is intended to be advisory and incorporates recommendations based on the experience of the authors as well as input gathered through a workshop held at the Network's Data and Information Systems national meeting.

2. Scope

This document provides guidance on reporting on study progress, oversight and monitoring during the life of the study, including producing reports, types of reports, potential risks and mitigations.

The guidance focuses on reports that are intended to be reused rather than ad hoc reports. This document does not cover tables and listings for statistical analysis (e.g. interim or final analyses), nor does it cover any narrative content.

3. 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.

4. Process

4.1. Types of reports

Throughout the life of a study, it is important to provide information about study progress and oversight to ensure the smooth running of a study and allow timely escalation and mitigation where issues are highlighted. See Appendix 1 for a non-exhaustive list of data reports.

4.2. Reports for target audiences

Specific data reporting outputs may be prepared for committees or other audiences. Multiple stakeholders (e.g. data manager, statistician, trial manager and clinician) may need to be involved in the agreement of the content and there should be clear documentation regarding which reports are provided for which purpose. Responsibilities for collating reports and circulating output should be made clear. It may be useful to develop standard reports

[1] Also known as DMEC, DSMB, IDSMC, IDMC etc.

(template output) to share with oversight committees so report content can be more efficiently agreed and programmed.

4.3. Agreeing content of reports

Reports should have a clear value and purpose, and teams should understand the resources required for specifying, developing and testing reports. It is important to properly consider and clearly define the specific requirements of a prospective report, with input from a multidisciplinary group including data management and programming roles as well as trial management and statistics, and others as relevant.

4.4. Process for producing reports

In order to reduce errors, improve quality, allow validation and save time, we recommend that reports are programmed in a reproducible way, rather than manipulating data manually. This should be carried out by trained individuals, using the software and methods recommended for that unit, and with validation processes agreed. Each unit should have a documented process covering agreement of content, format, production and validation of reports.

Reports should be built by suitably trained individuals, typically the data management team, software developer/programmer or statistician. Trial managers, QA/regulatory team and others (e.g. health economists) may be involved in running and collating reports but the technical aspect of the build and validation is likely to be performed by the more technical roles.

Reports may be based on real time data or data exported at a set point in time. Methods of producing, displaying and providing reports will differ depending on the functionality available. Reports may be interactive and allow for filtering (i.e. based on data ranges, sites etc.) by the person running the report.

Reports may be produced either with inbuilt reporting functionality within a CDMS or by using other applications. Other applications may interact with the data via:

- a download (“snapshot”) of the data at a point in time - typically used with a statistical software package (e.g. SAS, SPSS, R, Stata).

- a link to the live data, e.g. via Open Database Connectivity (ODBC) or an application programming interface (API) - typically used with data visualisation software (e.g. Power BI).

Reports may be available on demand within the CDMS or via a link to the live data, or on request. Alternatively, it may also be useful to schedule reports for set time points or at set frequencies.

When reports are distributed, consideration should be given to the audience and the purpose of the report. It may be useful to include a narrative that explains the content, or to ensure the report is presented by a member of the team who has a good understanding of the content of the report and underlying data.

4.5. Consideration of risks

It is important to consider potential risks of producing and sharing data reports, including the impact and likelihood of negative consequences (e.g. sometimes the potential effect of an incorrect/inappropriate/misinterpreted report output is catastrophic, sometimes negligible), and mitigate these where possible and appropriate.

Potential risks include:

- Unblinding (e.g. by including treatment allocation or presentation of results by arm).
- Unauthorised disclosure of personal or sensitive information.
- Providing incorrect information (due to incorrect programming).
- Incorrect interpretation caused by:
    - Incomplete data (e.g. due to delayed data entry).
    - Unclean data (e.g. outstanding data queries).
    - Lack of understanding of what the report is showing (e.g. around how particular calculations have been made).

It is important that risks regarding interpretation of data are highlighted before making important decisions about the running of a study (e.g. continuing or stopping recruitment, changing the sample size, requesting an extension to a funding period) based on report output.

4.6. Validation

Consideration should be given to how reports should be validated and how this validation should be documented. As with all computer system validation, a risk-based approach is recommended for making decisions around type and level of validation.

Common validation methods include:

- Code review.
- Manual review of output vs raw data.
- Use of test data.
- Independent generation of the intended output by different members of the team (double coding).

It's also important to consider and document, perhaps as part of the risk assessment, how often to validate these reports when using across different studies. If the report code is generic then it may be considered sufficient to validate when it is first developed and used in a particular study, then reused for other studies with minimal additional validation. On the other hand, if the report code needs to be amended significantly to use in a different study, it

will most likely be necessary to validate again per study, when the report is programmed / revised.

If it's agreed that changes are needed to a live report, then it's important to consider how much re-validation is needed to ensure that the output is still reliable.

If any changes are needed to the study database, the impact of any data structure changes on reports referencing this data should be assessed, and any necessary re-programming and/or re-validation performed.

5. Other considerations

To achieve efficiency and consistency, reduce the validation burden and avoid 'reinventing the wheel', it can be useful to maintain template code for reuse. Good commenting of code is useful for maintenance and reuse of template code. Using configurable elements for standard reports and having standard customisable templates applicable to different stakeholders can also be beneficial.

Use of a template database structure (e.g. standard field/variable names and coding) can help to maximise the benefits of this approach, especially for data that may be in a similar format across studies, e.g. consent, eligibility, and adverse events. Other key information such as treatment compliance may be difficult to completely standardise, but 're-use' should be considered wherever possible.

Other guidance documents will also cover the usefulness of adopting standards as they can save time at all stages: CRF design; database build; database validation; defining and testing validation rules; data reporting; and analysis. CTUs may develop their own internal standards and/or use industry standards (e.g. those put forward by the Clinical Data Interchange Standards Consortium (CDISC)).

Appendix 1 – Example report content

Note: This is not an exhaustive list, and the reports produced for an individual study should be based on study design and risks.

Recruitment

- Recruitment progress, e.g.
  - graph of actual vs target
  - Numbers screened, consented, eligible, recruited per site and totals
- Reasons for non-recruitment, including eligibility
- Randomised by intervention
  - It may be useful for some oversight committees to see randomised by intervention, however any risk of unblinding must be considered

Overview of participants

- Baseline characteristics
- Descriptive stats of key outcome measures at baseline
  - Keep in mind that including key outcome measures at follow up visits could inadvertently reveal (or appear to reveal) the results of the study, potentially leading to bias

Safety[2]

- Line listings of safety events of interest (e.g. SAEs) during the reporting period
- Cumulative summary tabulations of serious adverse events

Non-compliances/Protocol deviations

- Summaries (e.g. by type/site)
- Line listings with details

Study completion/discontinuation

- Early cessation of study treatment/active follow up/passive follow-up (ensure definitions are clear for each category – see Persevere guidance[3] for more information)

Data Quality & Completeness

- Query resolution rates
- Top queries (systematic issues)
- Data completion rates

[2] E2F Step 5 Note for guidance on development safety update report (europa.eu)

[3] https://perseverepriciples.org

- Summary of missing forms and fields
- Summary of overdue visits

Data Cleaning

- Line listings of data queries, missing forms etc[4]

Central monitoring[5]

- Consider guidance from the UKCRC Registered CTU Network Monitoring Group around choosing triggers and metrics[6].
- Consider triggers and checks to monitor these.
- Consider whether thresholds should be in a report or monitor compared to other sites/studies and expectations.

[4] Further information is provided within "Data Cleaning and Query Process" (to be published)

[5] See this paper for distinction between data cleaning & central monitoring: 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)

[6] https://ukcrc-ctu.org.uk/clinical-trial-monitoring/