**GCDMP**

**GCDMP©**

# Design and Development of Data Collection Instruments

Carolyn Famatiga-Fay*, Julie Filipenko[†], Evaldas Lebedys[‡], Derek Johnson[§], Rashida Rampurawala[‖] and Andrea Mulcahy[¶]

Data collection instruments are designed to record protocol-required data on each participant in a clinical study. One of the key clinical study documents, a case report form, is a type of collection instrument that allows the collection of accurate and complete study data. While this chapter primarily refers to case report form design and development including data definition, mapping, and guiding standards, most of the suggestions apply to other data collection tools. The regulatory basis for minimum standards and recommended best practices are also discussed.

## 1) Learning Objectives
After reading this chapter, the reader should understand:

· The purpose of and regulatory basis for case report forms (CRFs)
· Identification and definition of metadata
· Design, development, and maintenance of CRFs
· Establishing and maintaining libraries and standards

## 2) Introduction
In 2010, the Design and Development of Data Collection Instruments chapter was added to the Good Clinical Data Management Practices (GCDMP) published in 2013. It states, "If the data points specified in the protocol are not accurately collected, a meaningful analysis of the study's outcome will not be possible. Therefore, the design, development, and quality assurance processes of a CRF must receive the utmost attention."[1] This chapter carries the same message as in the previous version and further addresses the evolving practice of data collection in clinical trials. As stated in a recently published chapter of the GCDMP (2021) "Electronic Data Capture (EDC) Study Implementation and Start Up," in the first clinical studies, data were collected on paper forms. These were structured forms that served to ensure complete and consistent data collection for each study participant.[2]

In the development and design of instruments to capture appropriate data to support the analysis of a clinical study, it is crucial to know what data points to collect and how they will be collected. Designing CRFs plays a vital role in clinical studies, requires multidisciplinary team involvement, and warrants attention during the studies' development.

The International Conference on Harmonisation's Guideline for Good Clinical Practice E6(R3) defines the term "case report form" as, "A data acquisition tool designed to record protocol-required information to be reported by the investigator to the sponsor on each trial participant."[3] This chapter discusses considerations for CRF design, development, and quality assurance, including distinctions for studies using paper CRF and electronic data capture (EDC). It also provides further considerations, including maintaining a library of standards that will aid in ensuring that each CRF accurately and consistently captures data specified in a study protocol.

Because CRFs are related to numerous aspects of clinical data management (CDM), references are provided to other chapters of GCDMP that provide more in-depth information. While digital health technologies are not addressed in this chapter, it is important to consider them when planning a study's data flow, and whether some principles described in this chapter may apply. If data from digital health technologies are integrated as external data, refer to the "Integration of External Data" and "Electronic Data Capture (EDC) Study Implementation and Start-up" chapters.

## 3) Scope
This chapter focuses on the design and development of a CRF, a tool widely used in the collection of clinical data. From protocol review through development of standards, this chapter covers the routine study process in designing and

* Atara Biotherapeutics, US

† Graviton Bioscience, US

‡ SCOPE International, DE

§ Vaxcyte, US

‖ Independent, IN

¶ Alnylam Pharmaceuticals, US

Corresponding author: Carolyn Famatiga-Fay, MPH, CCDM (famatigafay@sbcglobal.net)

developing a CRF. In addition to adherence to regulatory guidelines and applicable standards, consideration is given to the following: **Data Identification** (protocol review), **CRF Design Practices** (formatting of how data should be collected, managing conditional logic, minimizing data redundancies, managing data privacy, handling language translations), **Development Considerations** (data mapping, edit checks, and data/system integration), **Libraries and Standards** (data element libraries and metadata, standards) and **Change Control and Versioning**. The chapter also provides information and guidance about relevant distinctions between paper versus electronic CRFs (eCRFs).

The focus on design and development of instruments in this chapter will be on collection of static rather than dynamic data. The distinction between the two will be addressed in this chapter.

For information about laboratory data and external data transfers, see the GCDMP chapter entitled "Integration of External Data". For more detailed information about EDC, see the 2021 GCDMP chapters entitled "Electronic Data Capture (EDC) Study Implementation and Start-up" and "Electronic Data Capture—Study Maintenance, Conduct, and Close Out". For more detailed information about different collection methods for patient-reported outcomes (PRO) data, see the GCDMP chapter entitled "Guidance for eCOA Development in Clinical Trials". CRF Completion Guidelines are referenced in this chapter and are addressed in detail in the 2021 GCDMP chapter entitled "CRF Completion Guidelines".

## 4) Minimum Standards

The **ICH E6(R3) Guideline for Good Clinical Practice** contains several sections relevant to CRFs.[3]

Section II of *Principles of ICH GCP* states, "The principles are intended to support efficient approaches to trial design and conduct… The use of technology in the conduct of clinical trials should be adapted to fit the participant characteristics and the particular trial design. This guideline is intended to be media neutral to enable the use of different technologies." It encourages the use of fit for purpose technologies to collect data.

Section 2.12 *Records* sets a requirement for the investigator to review and approve data: "The investigator should review and endorse the reported data at important milestones agreed upon with the sponsor."

Section 3.1 *Trial Design* provides considerations which are important for design of data collection instruments and promotes patient-centric and pragmatic trial concepts: "The sponsor should ensure that all aspects of the trial are operationally feasible and should avoid unnecessary complexity, procedures and data collection. Protocols, data acquisition tools and other operational documents should be fit for purpose, clear, concise and consistent. The sponsor should not place unnecessary burden on participants and investigators."

Section 3.4 *Qualification and Training* highlights the need for adequately qualified staff, including those responsible for CRF design: "The sponsor should utilise appropriately qualified individuals for the activities to which they are assigned (e.g., biostatisticians, clinical pharmacologists, physicians, data scientists/data managers, auditors and monitors) throughout the trial process."

Section 3.6 *Agreements* outlines that the sponsor is responsible for securing formal agreements with investigators, institutions, and, when applicable, service providers: "The sponsor should obtain the investigator's/institution's and, where applicable, service provider's agreements:

  (a)  To conduct the trial in accordance with the approved protocol and in compliance with GCP and applicable regulatory requirement(s)."

This requirement also implies the need to obtain appropriate licenses for proprietary assessment tools, such as validated questionnaires and rating scales, to ensure lawful and compliant use during the trial.

Section 3.10 *Quality Management* advocates a risk-based approach to quality management and emphasizes a need to integrate this approach in design and implementation processes: "Quality management includes the design and implementation of efficient clinical trial protocols, including tools and procedures for trial conduct (including for data collection and management), in order to ensure the protection of participants' rights, safety and well-being and the reliability of trial results. The sponsor should adopt a proportionate and risk-based approach to quality management, which involves incorporating quality into the design of the clinical trial (i.e., quality by design) and identifying those factors that are likely to have a meaningful impact on participants' rights, safety and well-being and the reliability of the results (i.e., critical to quality factors as described in ICH E8(R1))."

Section 3.16.1 *Data Handling* promotes design of fit for purpose data collection instruments and sets a requirement to follow protocol when designing data acquisition tools: "The sponsor should pre-specify data to be collected and the method of its collection in the protocol (see Appendix B)…The sponsor should ensure that data acquisition tools are fit for purpose and designed to capture the information required by the protocol. They should be validated and ready for use prior to their required use in the trial.". It sets a requirement for the sponsor to pursue investigator's approval of the reported data: "The sponsor should seek investigator endorsement of their reported data at predetermined important milestones."

Section 4.2.1 *Data Capture* indicates the importance of metadata: "Acquired data from any source, including data directly captured in a computerised system (e.g., data acquisition tool), should be accompanied by relevant metadata." It encourages further the integration of edit checks in the design of data collection instruments: "At the point of data capture, automated data validation checks to raise data queries should be considered as required based on risk, and their implementation should be controlled and documented."

Section 3.5.2 *Remote Data Collection Considerations in Appendix 2* emphasizes the need to consider specifics of decentralized trials when designing data collection instruments: "Remote data collection in clinical trials that incorporate decentralised and pragmatic elements (e.g., the use of remote visits and [Digital Health Technologies] (DHTs), such as wearables, or the extraction of data from [Electronic Health Records] EHRs) requires special attention to be paid to data security vulnerabilities, including cybersecurity and data privacy."

**Medicines and Healthcare products Regulatory Agency (MHRA) 'GXP' Data Integrity Guidance and Definitions** provides principles of data integrity, recommendations for establishing data criticality and inherent risk, designing systems and processes to assure data integrity.[4]

Section 2.6 states that study personnel "need to understand their data processes (as a lifecycle) to identify data with the greatest Good Practice (GXP) impact. From that, the identification of the most effective and efficient risk-based control and review of the data can be determined and implemented." In other words, CRF design should focus on the data that support study objectives and endpoints. Due to the cost and time related to data collection and processing, considerations should be given to collect the minimum data required. For example, the cost-benefit should be evaluated when considering collecting individual answers of each eligibility criteria versus collection of single confirmation that a participant is eligible.

Section 3.4 suggests that "Organizations are expected to implement, design and operate a documented system that provides an acceptable state of control based on the data integrity risk with supporting rationale. An example of a suitable approach is to perform a data integrity risk assessment (DIRA) where the processes that produce data or where data is obtained are mapped out and each of the formats and their controls are identified and the data criticality and inherent risks documented."

Section 5.1 sets the expectation on use of data integrity principles. It states, "Systems and processes should be designed in a way that facilitates compliance with the principles of data integrity."

Section 6.4 explains data integrity. It states, "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." It goes on to state that "The data should be collected and maintained in a secure manner, so that they are attributable, legible, contemporaneously recorded, original (or a true copy) and accurate."

Section 6.7 recommends that "Organizations should have an appropriate level of process understanding and technical knowledge of systems used for data collection and recording, including their capabilities, limitations and vulnerabilities." It emphasizes that "The selected method [of data collection and recording] should ensure that data of appropriate accuracy, completeness, content and meaning are collected and retained for their intended use."

The **General Principles of Software Validation; Final Guidance for Industry and FDA Staff (2002)** provides expectations on proper documentation of software used in clinical trials.[5]

Section 4.7 defines requirements for change control. It states, "Whenever software is changed, a validation analysis should be conducted not just for validation of the individual change, but also to determine the extent and impact of that change on the entire software system."

**The World Health Organization Technical Report Series No. 996 Annex 5 Guidance on good data and record management practices (2016)** consolidates existing normative principles, gives illustrative implementation guidance, and provides explanations as to what should be demonstrably implemented to achieve compliance. It focuses on principles that are implicit in existing WHO guidelines and that can affect data reliability and completeness, and can undermine the robustness of decision-making based upon those data if not robustly implemented.[6]

Section 4.3 emphasizes the need to apply the same principles for paper-based and electronic data collection and states, "The requirements for [General Data Protection Regulation] GDPR that assure robust control of data validity apply equally to paper and electronic data. Organizations subject to GXP should be fully aware that reverting from automated or computerized to manual or paper-based systems does not in itself remove the need for robust management controls."

Section 4.11 focuses on design of record-keeping methodologies and systems and states, "Record-keeping methodologies and systems, whether paper or electronic, should be designed in a way that encourages compliance with the principles of data integrity."

Section 4.12 provides further suggestions to achieve compliance with the principles of data integrity and recommends "controlling the issuance of blank paper templates for data recording of GXP activities so that all printed forms can be reconciled and accounted for."

Section 4.13 states, "Data and record media should be durable. For paper records, the ink should be indelible. Temperature-sensitive or photosensitive inks and other erasable inks should not be used. Paper should also not be temperature-sensitive, photosensitive or easily oxidizable. If this is not feasible or limited (as may be the case in printouts from legacy printers of balance and other instruments in quality control laboratories), then true or certified copies should be available until this equipment is retired or replaced."

Section 4.14 focuses on maintenance of record-keeping systems and states, "The systems implemented and maintained for both paper and electronic record-keeping should take account of scientific and technical progress. Systems, procedures and methodology used to record and store data should be periodically reviewed for effectiveness and updated as necessary."

Section 5.6 advocates the use of innovations and modern technologies. It states, "A data management programme developed and implemented upon the basis of sound

[quality risk management] QRM principles is expected to leverage existing technologies to their full potential. This in turn will streamline data processes in a manner that not only improves data management but also the business process efficiency and effectiveness, thereby reducing costs and facilitating continual improvement."

Section 11.4 provides recommendations for the good data process design and states that "Good data process design should consider, for each step of the data process, ensuring and enhancing controls, whenever possible, so that each step is:

· consistent;
· objective, independent and secure;
· simple and streamlined;
· well-defined and understood;
· automated;
· scientifically and statistically sound;
· properly documented according to GDPR."

Appendix 1 provides special risk management considerations for different aspects of data lifecycle management.

The section entitled *Special risk management considerations for review of original records* provides recommendations for data capture design. It states, "System design and the manner of data capture can significantly influence the ease with which data consistency can be assured. For example, and where applicable, the use of programmed edit checks or features such as drop-down lists, check boxes or branching of questions or data fields based on entries are useful in improving data consistency."

It goes on to specify, "The validity of the data capture process is fundamental to ensuring that high-quality data are produced. Where used, standard dictionaries and thesauruses, tables (e.g., units and scales) should be controlled."

**FDA Guidance for Industry Providing Regulatory Submissions in Electronic Format – Standardized Study Data** describes the requirements for an electronic submission of standardized clinical and nonclinical study data.[7]

Section C suggests, "When planning a study (including the design of case report forms, data management systems, and statistical analysis plans), the sponsor or applicant must determine which FDA-supported standards to use or request a waiver of those requirements…"

Section C.3 states, "The use of controlled terminology standards, also known as vocabularies, is an important component of study data standardization and is a critical component of achieving semantically interoperable data exchange… It is the expectation that sponsors or applicants will use the controlled terminologies maintained by external organizations as the standard."

The **FDA Study Data Technical Conformance Guide** provides specifications, recommendations, and general considerations on how to submit standardized study data using FDA-supported data standards.[8]

Section 1.2 states, "What data are collected and submitted is a decision that should be made based on scientific reasons, regulation requirements, and discussions with the review division. However, all study specific data necessary to evaluate the safety and efficacy of the medical product should be submitted in conformance with the standards currently supported by FDA and listed in the Catalog."

Section 4.1.4.6 advises, "The [annotated case report form] aCRF should include treatment assignment forms, when applicable, and should map each variable on the CRF to the corresponding variables in the datasets (or database). The aCRF should include the variable names and coding for each CRF item."

Section 6.1.1 states, "Controlled terminology standards are an important component of study data standardization and are a critical component of achieving semantically interoperable data exchange." It further specifies, "The analysis of study data is greatly facilitated by the use of controlled terms for clinical or scientific concepts that have standard, predefined meanings and representations … Controlled terminology is also useful when consistently applied across studies to facilitate integrated analyses (that are stratified by study) and cross-study comparative analyses (e.g., when greater statistical power is needed to detect important safety signals)."

Section 8.1.3 recommends that "An important component of a regulatory review is an understanding of the provenance of the data (i.e., traceability of the sponsor's results back to the CRF data) … Traceability can be enhanced when studies are prospectively designed to collect data using a standardized CRF, e.g., Clinical Data Acquisition Standards Harmonization (CDASH)."

The Appendix states, "Just as it is important to standardize the representation of data (e.g., M and F for male and female, respectively), it is equally important to standardize the metadata… In addition to standardizing the data and metadata, it is important to capture and represent relationships (also called associations) between data elements in a standard way. Relationships between data elements are critical to understand or interpret the data."

**PIC/S Guidance on Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments**[9]

Section 3.5 indicates, "The principles of data management and integrity apply equally to paper-based, computerized and hybrid systems and should not place any restraint upon the development or adoption of new concepts or technologies."

Section 5.3.4 states, "Not all data or processing steps have the same importance to product quality and patient safety. Risk management should be utilised to determine the importance of each data/processing step."

Section 5.5.3 recommends that "Risk assessments should focus on a business process (e.g., production, [quality control] QC), evaluate data flows and the methods of generating and processing data, and not just consider IT system functionality or complexity."

Section 8.4 sets "Expectations for the generation, distribution and control of records

· All documents should have a unique identification number (including the version number) and should be checked, approved, signed and dated.

- The document design should provide sufficient space for manual data entries.
- The document design should make it clear what data is to be provided in entries.
- Master copies should contain distinctive marking so to distinguish the master from a copy, e.g., use of coloured papers or inks so as to prevent inadvertent use.
- Updated versions should be distributed in a timely manner.
- An index of all authorised master documents, [Standard Operating Procedures] SOP's, forms, templates and records should be maintained within the pharmaceutical quality system."

Section 9.5 specifies expectations for data capture/entry: "Systems should be designed for the correct capture of data whether acquired through manual or automated means. All manual data entries of critical data should be verified, either by a second operator, or by a validated computerised means."

**European Commission Technical guidance on the format of the data fields of result-related information on clinical trials submitted in accordance with Article 57(2) of Regulation (EC) no 726/2004 and Article 41(2) of Regulation (EC) no 1901/2006** addresses the posting and publishing of result-related information in clinical trials. It defines the data fields to be reported for international harmonization and sets requirements for the data to be collected in clinical trials.[10]

**Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)** set principles relating to processing personal data and states that personal data shall be

- "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes"
- "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation')."[11]

**European Medicines Agency (EMA) Guideline on computerised systems and electronic data in clinical trials** describes general principles, key concepts, and expectations for computerised systems and electronic data for the data lifecycle.[12]

Section 4.3. *Data and metadata* indicates the importance of metadata for data integrity, and states that "The principles of data management and integrity apply equally to paper-based, computerized and hybrid systems and should not place any restraint upon the development or adoption of new concepts or technologies."

Section 4.4. *Source data* defines expectations relevant for direct data capture systems: "As a general principle, the source data should be processed as little as possible and as much as necessary."

Section 4.7. *Data capture* provides key principles for design of data capture systems.

- "The clinical trial protocol should specify data to be collected and the processes to capture them, including by whom, when and by which tools."
- "Data acquisition tools should be designed and/or configured or customised to capture all information required by the protocol and not more. Data fields should not be prepopulated or automatically filled in, unless these fields are not editable and are derived from already entered data (e.g. body surface area). The protocol should identify any data to be recorded directly in the data acquisition tools and identify them as source data."

Section 6.3. *Sign-off of data* suggests the consideration of a risk-based approach for the implementation of signatures.

- "The acceptable timing and frequency for the sign-off needs to be defined and justified for each trial by the sponsor and should be determined by the sponsor in a risk-based manner. The sponsor should consider trial specific risks and provide a rationale for the risk-based approach. Points of consideration are types of data entered, non-routine data, importance of data, data for analysis, length of the trial and the decision made by the sponsor based on the entered data, including the timing of such decisions. It is essential that data are confirmed prior to interim analysis and the final analysis, and that important data related to e.g. reporting of serious adverse events (SAEs), adjudication of important events and endpoint data, data and safety monitoring board (DSMB) review, are signed off in a timely manner. In addition, a timely review and sign-off of data that are entered directly into the eCRF as source is particularly important."

The **FDA Guidance for Industry, Electronic Source Data in Clinical Investigations** provides recommendations for capture, review, and retention of electronic source data. The guidance promotes capturing source data electronically and discusses ways of capturing it in the eCRF.[35]

Section 2.a *Direct Entry of Data Into the eCRF* explains that various data elements can be entered directly into the eCRF as source data:

- "Many data elements (e.g., blood pressure, weight, temperature, pill count, resolution of a symptom or sign) in a clinical investigation can be obtained at a study visit and can be entered directly into the eCRF by an authorized data originator. This direct entry of data can eliminate errors by not using a paper transcription step before entry into the eCRF. For these data elements, the eCRF is the source."

Section 3 *Data Element Identifiers* emphasizes importance of capturing data originator and data element identifiers in the eCRF:

· "The eCRF should include the capability to record who entered or generated the data and when it was entered or generated. Changes to the data must not obscure the original entry, and must record who made the change, when, and why. Data element identifiers should be attached to each data element as it is entered or transmitted by the originator into the eCRF."

Section 5 *Use of Electronic Prompts, Flags, and Data Quality Checks in the eCRF* highlights the importance of using electronic prompts, flags, and data quality checks in the eCRF to reduce data entry errors:

· "We encourage the use of electronic prompts, flags, and data quality checks in the eCRF to minimize errors and omissions during data entry. Prompts can be designed to alert the data originator to missing data, inconsistencies, inadmissible values (e.g., date out of range), and to request additional data where appropriate (e.g., by prompting a clinical investigator(s) to complete an adverse event report form triggered by a critical laboratory result)."

The **FDA Guidance for Industry, Conducting Clinical Trials With Decentralized Elements** provides recommendations regarding implementation of decentralized elements in clinical trials. The guidance discusses data collection in decentralized clinical trials (DCT).[20]

Section J. *Electronic Systems Used When Conducting DCTs* highlights DCT-specific considerations which are relevant for eCRF design and explains that "There are several ways local Health care Professionals (HCPs) can submit trial-related data for inclusion in clinical trial records, including but not limited to the following:

· An eCRF can be designed to allow local HCPs to enter trial-related data directly into the eCRF.
· Local HCPs can send forms or documents electronically by methods of secure data transfer (e.g., via secure email or fax) to investigators who are responsible for entering these trial-related data into the eCRF and retaining the trial-related records."

With these requirements in mind, in **Table 1** we state the following minimum standards for the design and development of data collection instruments.

**Table 1:** Minimum Standards.

1. Design CRFs that are fit for purpose, clear, concise and consistent to collect the data specified by the protocol and not more, with study objectives and endpoints in mind.[3,4,6,9,12]

2. Minimize and protect personal data collected.[3,11]

3. Ensure that CRFs have provision for investigator's signature as confirmation of data review; define and justify timing and frequency of sign-off based on the risk-based approach.[3,12]

4. Ensure that CRF design, development and approval is documented, and that change and version controls are in place.[5,6,9]

5. Ensure that only qualified staff design and develop CRFs.[3]

6. Ensure data collection instruments are available at the clinical sites prior to participant enrollment, and provide and document training and retraining of study personnel on data collection instruments.[3,12,25]

7. Verify CRFs that are based on rating instruments created by an independent source (eg, Health Status Questionnaire, Beck Depression Inventory, etc.), have been properly licensed for use and follow prescribed formatting or copyright requirements.[3]

8. Apply the same data integrity principles and controls to electronic and paper CRF and EDC.[4,6,9,12]

9. Ensure CRFs are designed and maintained in accordance with applicable regulations, guidelines, and standards which enhance traceability (ie, sponsor's results back to the CRF data) for regulatory review.[7,8,10]

10. Apply risk management and quality principles to all CRF design and development processes.[3,4,6,9]

11. Leverage innovations and modern technology in CRF design and development and do not hinder the development or adoption of new concepts or technologies.[6,9,12]

12. Use data element identifiers, audit trail, and electronic "prompts, flags, and data quality checks in the eCRF to minimize errors and omissions during data entry".[3,6,9,35]

13. Generate and maintain standardized metadata and annotated CRF (aCRF) or equivalent to accompany data collection tools.[3,8,12]

14. Design eCRFs to allow for direct data entry, when/where appropriate, with the source data processed minimally and only if necessary.[12,20,35]

15. Design data collection instruments considering data flow, data process, and technical specifications of the selected technology platforms, including their capabilities and limitations.[4,6,20]

16. Paper data collection instruments should provide sufficient space for manual data entries.[9]

17. Data fields should not be prepopulated, unless they are not editable and are derived from already entered data.[12]

## 5) Best Practices

Best practices, as stated in **Table 2**, are identified by both the review and by the authors of this chapter. Best practices do not have a strong requirement based in regulation or recommended approach based in guidance but do have supporting evidence, from either the literature or consensus of the writing group. As such, best practices, like all assertions in GCDMP chapters, have a literature citation where available and are always tagged with a roman numeral indicating the strength of evidence supporting the recommendation. Levels of Evidence are outlined in **Table 3**.

## 6) Data Identification and Data Definition

*FDA Guidance for Industry, Data Integrity and Compliance With Drug CGMP Questions and Answers*[19] distinguishes static and dynamic data: "static is used to indicate a fixed-data record such as a paper record or an electronic image, and dynamic means that the record format allows interaction between the user and the record content. For example, a dynamic chromatographic record may allow the user to change the baseline and reprocess chromatographic data so that the resulting peaks may appear smaller or larger. It also may allow the user to modify formulas or entries in a spreadsheet used to compute test results or other information such as calculated" field. This chapter focuses on design and development of instruments used to collect static data. Typically, dynamic data are captured and processed outside CRFs and thus, not within the scope of this chapter. While most of the considerations provided in this chapter apply to dynamic data, requirements specific to dynamic data should be taken into account when data collection instruments are used to capture dynamic data, eg, data collection instruments must be designed to preserve the dynamic record format.

### a) Protocol Review
#### i) Data Collection Tool Identification
It is imperative to identify the appropriate data collection tool using applicable regulatory guidelines,[16] eg, a CRF may not be the right tool to document and process protocol deviations or adjudications of centralized image reading. Systems limitation and/or data recurrence (certain procedures/events occur at every visit, others may vary visit to visit) may play a role in deciding where data should be collected.

#### ii) Protocol Version
CRF development should begin when a stable draft of study protocol becomes available or shortly after the final protocol has been approved, and it is critical that data collected in the final CRF reflects the approved version of the protocol.

#### iii) Required Data
CRF development is driven by a study protocol, which dictates data points required for assessment of the study's objectives and corresponding primary, secondary, and

**Table 2:** Best Practices.

1. Establish and maintain a library of standard CRF templates including annotations, data elements, conversions to other formats (e.g., CDASH to Study Data Tabulation Model (SDTM)), associated edit checks, and documents (CRF Completion Guidelines, participant diaries, etc.).[32,33] [III]

2. Use a multidisciplinary team to provide input into the CRF design and review processes. The study team, including CDM, data entry, statistical, safety and medical monitoring, regulatory, clinical, and clinical operations may be able to provide valuable perspectives to help optimize CRFs.[13,15,17,18,23] [III]

3. Keep the CRF's questions, prompts, and instructions clear, concise, and conformant to Clinical Data Interchange Standards Consortium (CDISC) – CDASH standards, where possible.[13,14] [III]

4. Design the CRF to follow the data flow from the perspective of the person completing it, taking into account the flow of study procedures.[13] [III]

5. Whenever possible, avoid redundant data points within the CRF. If redundant data collection is used to assess data validity, the measurements should be obtained through independent means.[13] [III]

6. Ensure and document training of clinical site personnel on the protocol, CRF completion guidelines, and data collection procedures.[13] [III]

7. Establish a procedure to manage translations and adequate quality control.[29] [III]

**Table 3:** Grading Criteria.

| Evidence Level | Evidence Grading Criteria |
|---|---|
| I | Large, controlled experiments, meta, or pooled analysis of controlled experiments, regulations or regulatory guidance |
| II | Small controlled experiments with unclear results |
| III | Reviews or syntheses of the empirical literature |
| IV | Observational studies with a comparison group |
| V | Observational studies, including demonstration projects and case studies with no control |

exploratory endpoints. The CRF development process is a vital quality control step, as it helps identify any data collection gaps within the protocol. Ensure that CRFs do not collect data that will not support research analysis required by the protocol. Including irrelevant data fields in the CRF may divert attention away from necessary data points and may impact the quality of data collected.

### iv) Schedule of Assessments
Start by reviewing the study matrix or schedule of assessments that summarizes assessments/procedures expected at each sequential visit or time point. In addition, careful review of data described outside of the schedule of assessments should be performed.[15] The protocol also specifies what data or results should be collected for each procedure/assessment/event in the study assessments and procedures section; this section may help determine data fields that need to be grouped together and placed/assigned to a specific CRF. Understanding how results (character values, decimal values, integer, etc.) and associated units are to be reported may help prevent updates to the CRFs.

### v) Statistical Input
Consult statisticians to ensure that all applicable key endpoints are collected in the CRF and how statistical features such as randomization, blinding, and variables (eg, collection of composite, surrogate, or categorical variables) may affect the CRF design.

## b) CRF Quality Checks Based on Protocol
The applicable stakeholders should perform the quality checks listed below once the CRF design is completed:

- All the applicable protocol specific procedures/events pertaining to study endpoints have corresponding forms and data collection fields in the CRF.
- All the procedures/events are collected in the appropriate visits/time points.
- The time points or window periods of all procedures/events are as per the protocol.
- The format of the data fields is in line with the protocol if specified. The data fields are free text only if specified in the protocol; otherwise, data fields should have protocol-specified options to select and enter the data in the CRF.
- The drop-down lists/codelists/data dictionaries associated with the data fields have all the options if specified in the protocol.
- The names/terminologies of the laboratory/test parameters are as per the protocol specifications and applicable standards, such as Common Terminology Criteria for Adverse Events (CTCAE), CDISC, etc. The applicable standards and/or libraries are used.
- The instructions on the CRF match the protocol.
- There is no duplication of data collection, ie, there is no overlap of the same data fields between CRFs.

## 7) CRF Design
### a) Human Form Interaction
CRF completion requires end user interaction and so it is important to consider how users intend to utilize the forms to complete their tasks and to ensure that these forms are user-friendly. Consider following the workflow study procedures and site requirements.[17] Authorized originators of clinical trial data and data collection workflow should be defined and must be taken into account when designing a CRF.[20]

### b) Graphics and Layout
CRF pages and fields should be arranged in a logical manner. Groups of data that are commonly collected together should have their associated fields placed together on the CRF.[17] For example, if in clinical practice, sites collect urine drug screens and urine pregnancy tests together, consider placing these on the same CRF page. Consideration should be given to the type and size of response, such as length of codelists or expected free text, if applicable. Note that some EDC systems may have limitations on graphics and layout. Design parameters such as colors, fonts, page placement, and appearance in different browsers or platforms affect user experience.

### c) Clarity and Ease of Use
Due to the nature of errors occurring during the initial entry into the CRF, design the CRF in a way that makes it clear and easy to use by end users. Doing so can help improve the overall quality of the data. Codelists and coded responses can contribute to ease of use by allowing users to check boxes or choose items instead of entering data into fields manually (refer to section e: Coded (Predefined) Responses). Avoid including too many fields on a single page, the crowded page may lead to cognitive burden.

CRF completion guidelines facilitate consistency in data collection, clarity, and ease of use. While some field-specific instructions can greatly support site personnel, overloading the forms with instructions should be avoided. For more information on CRF completion guidelines see the 2021 GCDMP chapter "CRF Completion Guidelines".

### d) Wording
Wording plays an important role in the ease of use and quality of the data collected. The intent of the questions should be clearly presented by using clear, concise language in the question or prompt.[18] Always avoid leading questions, and where possible, phrase questions in the positive to avoid the potential confusion that negatively stated questions can cause. For example, use "Did the participant complete this visit?" rather than "Was this visit not done?". It is important to consider the intended users' vernacular and avoid using any acronyms and abbreviations that may not be clear. Use plain language to ensure that users can easily read, concisely understand, and use the information provided. Follow a risk-based quality management approach and, when appropriate, plan risk mitigation measures such as readability testing, and the involvement of representatives of end users in user acceptance testing (UAT).

### e) Coded (Predefined) Responses
Implementing coded responses allows for data consistency and easy aggregation and facilitates analysis. A coded response is where the database restricts or populates the

possible responses that can be made, for example, Yes/No radio buttons (ie, codelists).

A set of coded responses and their order should be consistent throughout all forms. For example, if Yes/No questions are used throughout all CRFs, the order in which they are presented should be consistent throughout forms. Consideration should always be given to the end user and to their interaction with the form when creating coded responses. For example, sites may not be familiar with the CDISC controlled terminology and it is best to present the codelist with options that are easily understood by the users. Avoid using coded responses such as "check all that apply," as it leaves data open for ambiguity and interpretation. It is better in these cases to use an affirmative or negative response for each item.

When forms for direct data capture, ie, to capture source data, are designed, ensure that codelist options, their order, wording and formatting are not leading the user as it may introduce bias. Coded responses, like codelists, may limit the collection of accurate data and source records in direct data capture forms. For example, if a form collects symptoms a participant has experienced, the symptom codelist might not include all possible options, thus resulting in limited data.

### f) Conditional Logic

Conditional logic (also called "skip logic" or "branch logic") refers to displaying or hiding fields, or to skipping or generating forms based on a response on another field (parent field).[21] An answer to the parent field will determine whether additional child field/s should be answered. For example, the answer to the parent field "Did the participant have new lesions identified at this visit?" determines whether a new lesion details CRF is required. If "Yes" is selected, a new lesion details CRF will be made available for further data entry. If the answer is "No", the new lesion details CRF will not be made available for entry.

Clear instructions should be provided when additional questions need to be answered to avoid confusion. In designing paper CRFs, ensure that the additional CRFs needed to be made available based on the answer to a parent field are not placed in a remote location of the CRF casebook and that it is clearly marked and can easily be retrieved. It is important to provide clear instructions on the form itself and in the CRF completion guidelines on what is expected once site personnel or research participants answer the parent field.

In designing eCRFs, conditional logic is often managed by programming a dynamic function in the database. This functionality provides the ability to automatically show or hide fields, as well as generating or skipping forms once a response to the parent field is entered. This functionality should be tested during the database user acceptance testing process.

During revisions, any changes to the set/s of conditional logic fields must be tested to ensure that the parent field still exists and that the related additional fields are still valid. Deletion or alteration of the parent field may result in deleting or altering child fields.

### g) Minimizing Redundancy

#### i) Data Redundancy

Data redundancy occurs when duplicate data is entered in more than one place and can result in one or more of the following:

- same data entered on multiple CRF pages and/or multiple database systems;
- data entered not matching expected value (ie, value entered reflects a different unit of measurement than expected);
- additional edit checks;
- an increase in resource needs for data entry, data reconciliation, and query resolution; and
- confusion over which of the data elements should be the correct source for analyses.

Because of these potential consequences, data redundancy should be minimized or eliminated in the CRF design. For studies of adult populations, for example, plan to collect height once instead of collecting height at each visit, as it will most likely not change.

#### ii) Derived Fields

Avoid designing eCRFs in which sites are expected to perform calculations that are subject to error even if raw data needed for the calculations are being collected and entered on the eCRF. For example, Body Mass Index (BMI) can be programmatically calculated based on raw data (height and weight) and should remain as a derived field rather than requiring sites to calculate and populate the field.

#### iii) Intentional Redundancy

In a few instances, seemingly redundant data may be collected to assess data validity. An example is confirming a pregnancy test result by collecting it twice. In these instances, the same data should be obtained using different measurements or methods. For example, both urine and blood samples can be used to validate a pregnancy test result for the same time point and if both test results are the same then data is deemed valid.

#### iv) Log Forms

Collection of multiple events or records, such as adverse events or concomitant medications, could result in unintentional data redundancy. Although log forms allow for multiple rows in one form, thus allowing duplicates to be spotted, they do not eliminate the problem completely. In the case of the collection of multiple events or records with log forms, edit checks should be employed to detect overlaps in dates rather than relying on visually detecting duplicates.

#### v) Reconciliation

When using multiple data collection tools, it is important to reconcile data. Prior to finalizing the CRF, check if there will be other data collection tools and whether external data will be integrated with the clinical database. If any duplicate data point is collected in those systems, evaluate whether it can be limited to one data collection tool. Follow a risk-based approach when evaluating the need to collect data points in multiple databases. If it cannot

be limited to collecting once, ensure that appropriate edit checks are programmed and/or reconciliation process are in place.[22]

### h) Paper CRF Distinctions
While use of eCRF has become prevalent, paper CRFs may still be used when an EDC system is impractical.[23] Although there may be disadvantages in using and designing paper CRFs, there are instances where paper CRFs may be a more appropriate option for collecting data. There are certain factors that should be considered, such as situations in which training personnel to use an EDC system would be too costly or disruptive to routine practice. Apply the same quality control and data integrity to the design of both paper CRFs and eCRFs. Additional consideration must be given to CRF printing quality, layout, CRF Completion Guidelines, and transcription method.

#### i) Printing
A paper CRF must be printed in a way that reduces the potential for missing data due to questions being overlooked as the form is being filled out. For example, paper CRFs should be printed single-sided and should use a clearly legible font style and size. Including too many data fields in a single CRF can lead to questions being skipped, because the page may become too crowded for the eye to discern different items easily. The layout of questions on a paper CRF should support good ergonomics in how data items are displayed and the header section of the CRF should be clearly defined with the applicable form name.[23] Printing of paper CRFs should be accompanied by adequate quality control to ensure data integrity. When copies of original CRFs are used to collect data, the risks should be assessed and appropriate mitigation measures implemented because photocopies and scans can obscure images and text or can truncate sections of a page. CRF printing is one of the common services that are outsourced to a third-party vendor. For more information on how to evaluate and select vendors, see the 2021 GCDMP chapter "Vendor Selection and Management."

Refer to *Appendix A − CRF Printing* within this chapter for CRF printing guidance, extracted from the 2013 GCDMP Chapter "CRF Printing and Vendor Selection" − *May 2007*.[1]

#### ii) Administrative and Tracking Design Elements
Unlike eCRFs, paper CRFs should contain additional design elements for administrative and tracking purposes. For example, each CRF page should contain both the page number and the total number of expected pages to reduce the likelihood of missing data. Each CRF page should also contain identifying information linking the data collected to the correct protocol, site, research participant, and time point.

Instructions for language, format, standard units, and coded responses should be provided in the CRF Completion Guidelines, with applicable guidance also printed on the CRF page itself. For example, wherever dates are recorded on a paper CRF, the expected date format should be clearly stated, especially in studies that span multiple countries or geographic regions. Ideally, dates should be entered according to the CDASH standard format of using a 3-letter abbreviation for the month (DD-MON-YYYY), which avoids the potential ambiguity of different date formats. It is also important to consider how partial dates should be entered if the exact date is not known. If times are requested, they should be recorded using the 24-hour clock and hh:mm:ss format, with the appropriate level of precision needed for that field. Unit of measure (eg, kilograms or pounds, centimeters or inches) should also be clearly identified.

### i) Optical Character Recognition (OCR)
Manual entry of CRFs or processing via optical field or character recognition (OCR) scanners must be accounted for when designing the paper CRF. Some studies suggest that OCR systems may have unacceptable error rates for transcribing data and require significant time to validate. For databases using simple questionnaires, the OCR method of transcription may be feasible, although more complex or open-text data fields may be less suited for this technology.[24]

### j) eCRF Distinctions
#### i) eCRF Advantages
EDC systems have tools such as eCRFs or electronic patient-reported outcomes (ePRO). Per the FDA Guidance for Industry, Computerized Systems used in Clinical Investigations, "There is an increasing use of computerized systems in clinical trials to generate and maintain source data and source documentation on each clinical trial subject. Such electronic source data and source documentation must meet the same fundamental elements of data quality … that are expected of paper records."[25] eCRFs may have advantages over paper CRFs in terms of time spent on data entry and increased data quality.[24] eCRFs can also simplify the collection and tracking of data across multiple study centers, countries, or geographical regions.[26] Issues with data attributability and legibility are eliminated when moving from a paper-based to electronic instrument as the latter has an electronic audit trail that automatically captures users associated with a data point, date, time, and reasons for change. Specifically, for ePRO data, use of an electronic instrument such as an e-diary ensures increased protocol compliance, data validity, measurement reliability, and auditable quality.[27]

Use of electronic instruments reduces many of the risk mitigations needed to maintain data cleanliness or validity described in the *Paper CRF Distinctions* section h) above. For example, the eCRF has a drop-down or pop-up calendar functionality that eliminates the possibility of inconsistent or incorrect date formats. The eCRF can also provide a clear option for partial or unknown dates or times to be entered.

eCRF enables single entry of identifiers such as site number, participant number, and time point when first initializing the participant's casebook in the system. This helps avoid potential errors associated with incorrect participant or site identifiers on a CRF.

Edit checks programmed within the eCRF help validate data at the point of data entry. Paper CRFs, on the other hand, cannot flag errors or validate data at the point of data entry as edit checks are not performed until data is transcribed into the clinical data management system, removing the possibility of real-time data cleaning.

To further limit errors, since eCRFs can also be prone to errors in data entry and transcription from paper source documents,[28] additional validation and testing will be required when designing eCRFs to ensure that they function as expected, and meet regulatory guidelines for accuracy, traceability, and attributability. This would incur additional costs that would not otherwise be observed when designing paper CRFs.

### ii) eCRF Design Considerations

Just like paper CRFs, eCRFs should be designed with the end-user interface in mind; eg, avoiding crowding the screen with unnecessary text or data fields.[23] In contrast to a paper CRF, which can be organized into one group of questions or multiple sub-groupings, division of the eCRF into logical subsections is preferred, in part to allow data to be saved before moving on to another subsection. Designing the eCRF using subsections reduces the risk of losing unsaved data because of a system or server issue during data entry.[23] eCRFs offer the capability to tab through fields in a prescribed sequence, which can help minimize the chances of a question being overlooked.

According to ICH E6(R3) Good Clinical Practice: Integrated Addendum, "The sponsor should ensure that all aspects of the trial are operationally feasible and should avoid unnecessary complexity, procedures, and data collection. Protocols, case report forms (CRFs), and other operational documents should be clear, concise, and consistent."[3] Complex branching logic should be avoided as it may introduce unnecessary restrictions that can affect data quality. Risk assessment should evaluate the bias, which complex eCRF design may introduce, and adequate mitigation measures should be planned.

### iii) Integration Facilitation

Well-designed eCRFs within EDC facilitate integration of the clinical database with other data capture systems, such as the clinical trial management system (CTMS) and safety database, reducing the likelihood of inconsistent data across multiple data sources.[22] If the clinical data is to be integrated with other systems, the data manager should consider which data fields are fed into the EDC from an external source, which data fields may reside in both EDC and other systems, and whether centralized dictionaries and codelists should be used across systems.[22] Designing the eCRF to prepopulate data points from other systems may result in considerable timesaving effects.[24] For example, integrating Medical History data from EHRs into eCRF would save time as site staff would not have to enter data, the site monitor would not have to verify it, and the data manager would not have to design edit checks to check its validity. Similarly, integrating Laboratory data from labs into eCRF would relieve a burden of reconciliation and querying.

For more information about eCRF design within EDC, including dynamic functionality of forms, fields and time points, see the 2021 GCDMP chapter entitled "Electronic Data Capture (EDC) Study Implementation and Start-up."[2]

### k) Clinical Outcome Assessment (COA) Forms

A clinical outcome assessment (COA) is a measure that describes or reflects how a study participant feels, functions or survives and may be completed by a participant (patient-reported outcome), non-health professional observer (observer-reported outcome (ObsRO)), or a trained health professional (ClinRO). Types of COA include:

· Patient-reported outcome: data that are directly reported by research participants without alteration or interpretation by clinician or others, or when completed electronically (ePRO).
· Observer-reported outcome (ObsRO): a measurement based on a report of observable signs, events or behaviors by someone other than the study participant or a health professional (eg, parent, caregiver, etc.) and by clinician, respectively.
· Clinician-reported outcome (ClinRO): a measurement based on a report that comes from a trained health-care professional after observation of a participant's condition.

This type of data is crucial to studies that attempt to quantify research participants' subjective experiences, such as pain intensity or quality of life, using measures that include rating scales, questionnaires, interviews, and counts of events.

If the COA data entered in the data collection instrument is based on a scale or tool from an independent source (eg, Health Status Questionnaire, Beck Depression Inventory); then the validity of that instrument must be maintained. If any changes in content or format are necessary, then an independent source should be consulted to ensure that the validity of the tool has not been compromised by the changes. Documentation of all changes as well as maintaining the validation of the tool after changes have been made. Also, confirm that all necessary licensing and copyright requirements have been satisfied.

When developing a new data collection instrument to capture COA data, the same concepts and best practices for designing eCRFs and paper CRFs apply. For example, consider whether the questions are meant to elicit a structured or an open-ended response, and whether the text responses need to be converted to coded variables in the database or not.[15] For more information about COA and eCOA data collection, see the GCDMP chapter entitled "Guidance for eCOA Development in Clinical Trials."

### l) Translations

Clinical studies conducted globally and in countries with multiple languages often require translations of data collection instruments to accommodate the language that is primarily used in a participating country. The meaning of the data collected, regardless of language, should hold the same definition and therefore, interpretation. The

list of translated data collection instruments and quality control measures should be in the Data Management Plan or equivalent study document. Procedures should be in place for change control and versioning of translations. Provisioning of access to applicable translations should be controlled and respective documentation should be maintained.

Translation should be 1) clear and easily understood, 2) expressed in language in common use, and 3) conceptually equivalent to the original.[29] CRFs translated into multiple languages (including Braille for the visually impaired) should be carefully reviewed to ensure the translations are truly equivalent. One method of ensuring equivalency, for example, would be for one party to translate the CRF into the target language and then for a second party translate this back to the source language and compare the results against the original document.

Setup of multilingual CRFs can be resource intensive and may add complexity to the management of collected data, especially when free text data are collected in different languages. CRFs should be designed and translations implemented with consideration given to the way collected data will be managed, eg, specific fields for translations and quality control of translations may need to be implemented. When the same language is spoken in more than one country, cultural and other differences need to be taken into account when implementing translations.

### m) Data Privacy
Data privacy must be maintained to protect the confidentiality of personal data of the study participants.[3] The CRF must avoid collecting data that could lead to direct or indirect identification of the study participant.[11] It is imperative that data privacy is handled in compliance with local regulations and the EU GDPR. In designing the CRF, a participant identifier should maintain confidentiality of the personal information of the participant.

Personal data should not be collected in the CRF, unless required by the protocol. Refer to the company SOPs and to GDPR for details on what is defined as personal data. Certain data fields, whose content could be considered a confidentiality breach, may be allowed to be collected partially in certain locations. For example, in some regions only the year of birth may be allowed to be collected for date of birth whereas in others, the entire date of birth can be collected.[14] When designing forms such as informed consent, consider including data fields that collect the consent of study participants for using their data from previously collected samples or for future research analysis. CRF prompts and instructions must be clear and guide site personnel on what and how instructions should be conveyed to study participants, eg, vague and overly broad language should be avoided.[3] It is equally important to collect the consent withdrawal of the study participants. The study participant may choose not to share certain types of data (eg, biospecimens, imaging, genomics, etc.) and the CRF should reflect it. For more information about privacy issues in clinical research, see the 2013 GCDMP chapter entitled "Data Privacy".

### n) Study Disruptions
In exceptional circumstances, such as public health crises or natural disasters, which may result in difficulty collecting data as planned, it is important to refer to regulatory agencies (eg, the FDA) for guidance on how to properly document data collection procedures that may have been disrupted. CRF design should account for missed or remote visits/assessments, exposure changes, site transfers, etc.

For example, in response to the significant disruptions caused by the COVID-19 pandemic, CDISC swiftly formed a task force to address the emerging challenges in clinical trials. Recognizing the urgent need for standardized approaches to manage impact of pandemic on ongoing and new studies, CDISC published recommendations.[30]

## 8) Development Considerations
### a) Data Mapping
A data element is a unit of data for which the definition, identification, representation, and permission values are specified through a set of attributes: name, type, caption presented to users, detailed description, and basic validation information (eg, range of values).[23]

Data field names are created and are attributed to a specific data point on a CRF. aCRFs are blank CRFs with the assigned field names mapped to the data point. The annotations are used to identify the fields not by the question text on the CRF but by an assigned name that represents a particular question on the CRF. These names are also known as variable names. These annotations tend to be short and are used for programming purposes. They facilitate datapoint tracking and help with derivation calculations and data analysis and interoperability.

In-file data that describe the attributes of other data, and provide context and meaning are a form of metadata. Typically, these are data that describe the structure, data elements, inter-relationships and other characteristics of data, eg, audit trails. When documenting these attributes, include complete codelists, and the order of the coded responses, labels, prompts and instructions as they appear in the data collection instruments. Metadata also permit data to be attributable to an individual (or if automatically generated, to the original data source). Metadata form an integral part of the original record. Without the context provided by metadata the data has no meaning.[9]

Data mapping begins with the creation of aCRFs and/or the creation of CRF specifications. It is recommended that once annotations are finalized, they are not changed without consideration of downstream impact on workflows such as programmed reports and calculations/derivations. Accurate data mapping is crucial in the ability to program derivations, organizing data in the format of reports, and eventually accurately representing data elements that will be used for final data analysis. When the CRF is revised, careful review of existing annotations should be performed if fields are updated, added, or removed.

Achieving traceability is important for regulatory submissions (ie, allowing reviewers to track any changes to a data point from the time it was entered in the CRF, including reasons for changes made, until final submitted

data). There may be instances in which a data field is represented differently in a dataset used for analysis, depending on the standards the sponsor organization is using. For example, per CDASH standards, Respiratory Assessment date and time fields may be named as "REDAT" and "RETIM" on the aCRF. When re-mapped to a SDTM format, these fields are combined into one SDTM field as "REDTC". The mapping from one format to another should be specified in these instances, based on the standards used, which will aid in ensuring traceability.

### b) Edit Checks

Regardless of how well CRFs are designed, edit checks should be programmed into the database or clinical data management system (CDMS). Edit checks are intended to ensure data integrity and improve data quality by bringing attention to data that are out of the expected range, inconsistent, illogical, or discrepant. When data meets the predefined criteria of an edit check, a flag or warning known as a query is generated that notifies personnel that the data point should be carefully examined to ensure its accuracy.

For more information about edit checks, see the 2013 GCDMP chapter entitled "Edit Check Design Principles"

### c) Investigator Signatures

Per ICH E6(R3) Good Clinical Practice, "The investigator should ensure the accuracy, completeness, legibility, and timeliness of the data reported to the sponsor in the CRFs and in all required reports."[3] Those data should be reviewed and signed by investigators on an ongoing basis. The investigator's signature is considered to be the documented confirmation that the data entered in the CRF and submitted to the sponsor are attributable, legible, original, accurate, and complete and contemporaneous. An investigator must have oversight of the study conduct at the site and documented timely data review is a way to demonstrate this. The sponsor should decide on the frequency and time points for CRF sign-off using the risk-based approach that takes into account the study duration for each participant, the criticality of data, the data analysis time points, etc. Considerations should be given to inclusion of signature manifestation and respective regulatory requirements.

### d) Data/System Integration & Other Technology Considerations

#### i) System Integration

Data flow, data exchange and system integrations should be considered at the study start-up and defined in the Data Management Plan with an overview of all data collection instruments and systems used to capture data required by the protocol. should be Refer to the "Electronic Data Capture (EDC) Study Implementation and Start-Up" and "Integration of External Data" chapters for more information on data that are transferred to a clinical database, such as data from a Randomization and Trial Supply Management (RTSM), a diagnostic imaging device, or an ePRO/eCOA device. In cases where integrated data must be displayed in the eCRF (eg, randomization data), design the form clearly, indicating which fields are integrated and do not require data entry in eCRF. The design of data collection tools should leverage existing scientific and technical progress, modern technologies and innovations, and be designed in such a way that does not hinder future technical advancement.

#### ii) Direct Data Capture (DDC)

Data collection is evolving toward the direct use of eSource data (eg, EHRs, wearable devices, etc.) using the Health Level Seven (HL7®) Fast Healthcare Interoperability Resources (FHIR®) standard, which is designed to address the limitations of older data exchange standards, such as HL7® Clinical Document Architecture (CDA®) or the Integrating the Healthcare Enterprise (IHE) Retrieve Form for Data Capture (RFD) standards. All these standards aim to streamline data acquisition by eliminating steps (such as manual transcription between electronic systems and source data verification by the monitor during a site visit) currently needed to transport clinical data from a physician's participant medical charts to a study's clinical database. Because every data processing step introduces the potential for error, HL7® FHIR® may soon be a huge contributor to improving data quality while also reducing study costs and timelines.

Data from clinical assessments is typically captured first in paper or in electronic source records and then transcribed into a CRF. However, direct data capture (DDC), ie, direct entry into a CRF with no prior entry into the source records, may be accepted when defined and approved in the study protocol. In this setting, "eSource DDC" refers to an electronic system that allows direct entry of source data, and some of these data are defined as CRF data collected for clinical study purposes. When eSource DDC is used, data transcription from one place to another, eg, from source to CRF, can be avoided. Where specified in the protocol, the eSource DDC may be the original point of recording specified information. A typical example would be investigator rating scales and detailed recording of multiple blood sampling times; they are not used in normal clinical practice but are collected per specific study requirements. For such data, the direct recording into eSource rather than an initial recording in a medical record that is later transcribed into an eCRF is likely to improve data quality. While the FDA acknowledges advantages of DDC in "Guidance for Industry, Electronic Source Data in Clinical Investigations"[35] worldwide regulatory acceptability of DDC varies; conformance with regional and national legislation and data protection requirements, therefore, need to be assessed.

Guidance and recommendations given in this chapter, such as clarity of use, minimizing redundancies, data privacy, etc. should be taken into account when planning and implementing integration and/or DDC.

## 9) Libraries and Standards
### a) Data Element Libraries and Metadata

Use of libraries and standards can greatly decrease both the cost and time of CRF development, help harmonization and integration and analysis of data across different systems and studies, and increase interoperability, defined by Petavy

as "the ability of different information technology systems and software applications to communicate, exchange data, and use the information that has been exchanged"[31] Data standards also facilitate community engagement, data sharing, transparency, and improved policymaking.[31]

Some organizations create and maintain a library of standard CRF templates, including annotations, data elements and associated codelists and edit checks, which allow CRFs to be easily modified to meet the needs of each individual study and enable robust maintenance and documentation of metadata. Any conversion to other formats (eg, CDASH to SDTM) should also be considered as part of the standards library to be developed.[32,33]

- CDISC eCRF Portal consists of ready-to-use, CDASH-compliant, annotated eCRFs, available in PDF, HTML and XML, to use as is or import to an EDC system for customization.[36]
- The Medical Data Models (MDM) is a registered European research infrastructure based mainly on CRFs.[34]

### b) Standards
In addition to standardized CRFs, other standards that might impact CRF design come from various sources.

#### i) CDISC
CDISC first released the CDASH standard in October 2008, with the intention to standardize CRF data collection fields. CDASH provides a set of data collection fields and controlled terminology that are divided into domains; it is designed to be applicable to clinical studies regardless of therapeutic area or phase of development. For more information about CDASH and other standards that affect CDM, see the 2013 GCDMP chapter entitled "Data Management Standards in Clinical Research." At the end of 2016, the FDA announced that some of the CDISC standards are now mandatory for regulatory submissions.[7,37]

#### ii) Medical Subject Headings
CRFs may be coded using Medical Subject Headings, developed by the National Library of Medicine (MeSH; https://www.nlm.nih.gov/mesh/meshhome.html); ie, terms for the indication under study in the trial in which the CRF was used.[33]

#### iii) Regulatory Standards
Regulatory standards, such as the GDPR may have an impact on CRF design, particularly concerning data privacy or CRFs that are translated into multiple languages.

#### iv) Data Standards Initiatives
Several initiatives launched within the FDA's Center for Drug Evaluation and Research (CDER) are aimed at the development of data standards and at the management of clinical data across specific therapeutic areas.

*PhUSE*: "At the March 2012 FDA/Pharmaceutical Users Software Exchange (PhUSE) Annual Computational Science Symposium, working groups-initiated discussions on validating data, improving data quality, standardizing

data within the site selection process, exploring the challenges of integrating and converting data across studies, identifying standards implementation issues with the Clinical Data Interchange Standards Consortium (CDISC) data models, developing standard scripts for analysis and programming, and creating the nonclinical road map and its impact on implementation."[38]

*Critical Path*: "The Critical Path Initiative introduced in 2004 has encouraged industry, academia, and government agencies to develop public–private partnerships (consortia) in order to collaborate and share information, technology, and expertise to bridge the gap between scientific discoveries and their translation into innovative medical therapies. Two examples of partnerships that are incorporating data standards into their efforts are the Coalition Against Major Diseases (CAMD) consortium and the Analgesic Clinical Trial Translations Innovations, Opportunities, and Networks (ACTTION) public–private partnership. CAMD, a consortium convened through the Critical Path Institute, worked with CDISC to develop a user guide for standard data elements for Alzheimer's disease (AD) and utilized these standards to create a database of AD studies."[38]

#### v) EDC
The design of CRFs in studies using EDC systems is often influenced by the technical standards and limitations of the chosen EDC platform. For example, in theory, Collection Date field LBDAT could be used for all laboratory (eg, Chemistry and Hematology) forms. However, in practice, having the same field in two different forms may not work with EDC's edit check design, requiring distinct field names (eg, HEM_LBDAT and CHEM_LBDAT).

#### vi) International System of Units
A conversion factor table to standardize conversion of conventional units to the International System of Units (SI, abbreviated from French Système international d'unités) should be used.[39]

## 10) Change Management
The CRFs should be managed as controlled documents and, as such, should be subject to change control. When CRFs undergo changes, the key principles of change control need to be followed to ensure traceability and documented evidence of what was done, when, by whom, why, and how. CRF change management should be governed by organizational or study-specific procedures and cover the following aspects:

### a) Impact of changes
Before CRF changes are initiated, the impact of changes should be assessed on

- Data; ie, consistency of data collected prior to and after change. The impact on the captured data should be assessed with care and adequate controls should be considered to ensure that collected data is not affected, is accessible, and stays under control of the investigator.

- Data systems and system integration.
- End users, including additional user manuals, training needs, system downtime, data re-entry, or migration, etc.

Changes should be implemented with respect to impact assessment and risk–benefit considerations.

### b) Quality control and approval

CRFs should reflect the current version of the protocol and need to be updated when protocol amendments are released. Ideally, the team involved in CRF design should be involved in the development of the protocol amendments. Changes need to be implemented in a controlled manner, including reviews by subject matter experts, validation, and UAT as appropriate. CRF Completion Guidelines should be updated and reviewed in line with changes made in CRFs. Revision or version history, ie, the changes made with the rationale of changes and references to affected CRF versions, should be maintained.

### c) Versioning conventions and revision status identification

All CRFs should have unique identification that indicates version and revision status, ie, draft/final. If only single forms are affected by the change, the new version of single forms can be released with clear version identification. Clear version identification, including study version and document version, should be visible on each CRF. The metadata of clinical data records should have association to the CRF version used.

### d) Release, distribution, control, and management of access to the relevant versions

Prior to new versions becoming effective or released, consideration should be given to any regulatory or local requirements, eg, when appropriate, regulatory authorities should approve updated CRFs before new versions are used to collect data. Updated versions should be distributed in a timely manner. All study personnel must be made aware about the changes before updated CRFs become effective.

### e) Restricted access to obsolete versions

Controls should be implemented to ensure that the current CRF versions are used. Previous versions should be accessible in read-only mode.

## 11) Recommended SOPs

Section 5.0.1 of ICH E6(R3) states that, "During protocol development the sponsor should identify those processes and data that are critical to ensure human subject protection and the reliability of trial results."[3] This implies that organizations should map out the processes involved in study design, start-up, conduct, and closeout, and they should make explicit decisions about which are considered to impact human subject protection and the reliability of trial results. Organizational processes may be partitioned differently, which may lead to different scopes and titles for SOPs. The following is presented as a list of processes commonly considered to impact human subject protection and the reliability of trial results. Organizations may differ as to how these processes are covered in SOPs. The recommendation of CRF Design and Development and CRF Printing Specifications SOPs is based on consensus of the writing group, including the GCDMP Executive Committee and opinion papers.

- CRF Design and Development
- CRF Quality Assurance
- CRF Approval
- CRF Version Control
- CRF Printing Specification
- Vendor Selection and Management (Title 21 CFR 312.52,[40] ICH E6 R3 5.0[3])

## 12) Literature Review

This revision is based on a systematic review of the peer-reviewed literature indexed for retrieval. The goals of this literature review were firstly to identify published research results and reports of evaluation of new methods regarding Design and Development of Data Collection Instruments and secondly to identify, evaluate, and summarize evidence capable of informing the practice of Design and Development of Data Collection Instruments.

The following PubMed query was used:

*("data collection form" OR "CRF" OR "Case Report Form" OR "medical record abstraction form" OR "chart review form") AND ("design" OR "development") AND ("clinical trial" OR "clinical trials" OR "clinical study" OR "clinical studies" OR registry OR registries OR "observational study" OR "interventional study" OR "phase 1 study" OR "phase 2 study" OR "phase 3 study" OR "phase 4 study" OR "phase I study" OR "phase II study" OR "phase III study" OR "phase IV study" OR "first in man" OR "clinical research" OR "device study" OR "interventional trial" OR "phase 1 trial" OR "phase 2 trial" OR "phase 3 trial" OR "phase 4 trial" OR "phase I trial" OR "phase II trial" OR "phase III trial" OR "phase IV trial" OR "randomized clinical trial")*

The search query was customized for, and executed on, the following databases: PubMed (267 results); EMBASE (490 results); Science Citation Index/Web of Science (354 results); Association for Computing Machinery (ACM) Guide to the Computing Literature (10 results). A total of 1153 works were identified through the searches. The searches were conducted on May 17, 2021. Search results were consolidated to obtain a list of 746 distinct articles. Because this was the first review for this chapter, the searches were not restricted to any time range. Literature review and screening details are included in **Figure 1**.

Two reviewers used inclusion criteria to screen all abstracts. Disagreements were adjudicated by the writing group. Of the 57 articles that met the inclusion criteria and were selected for review, 28 were not accessible. The selected articles were read by the writing group.
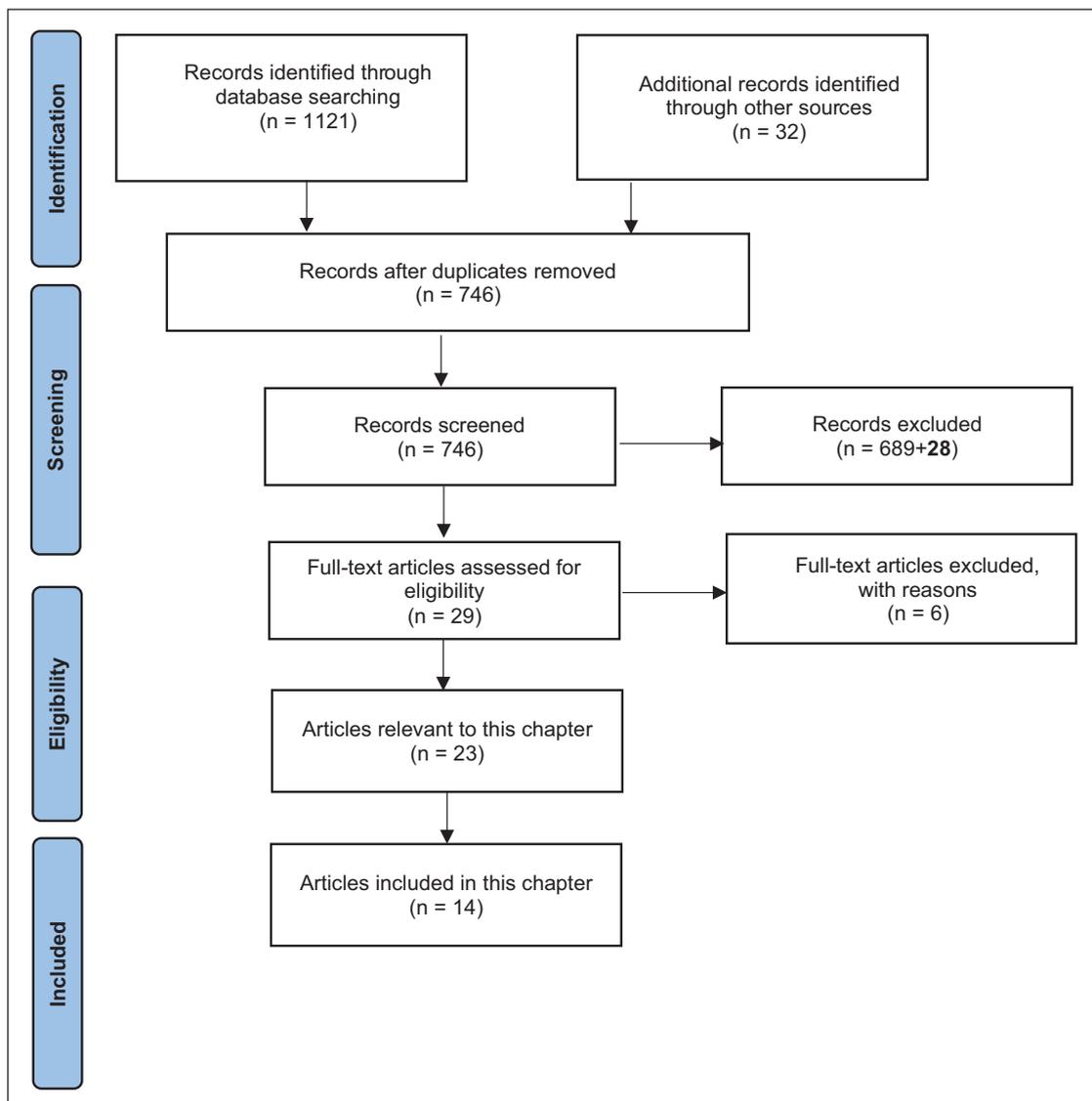
**Figure 1:** Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) for the Design and Development of Data Collection Instruments.

Each of the 29 articles was read for mention of explicit practice recommendations or research results informing practice. A total of 23 articles were deemed relevant to this chapter and six were excluded by the full text review as not relevant. Of the 23 relevant articles, 14 were identified as informative of practice and 9 were relevant but not informative. Fourteen articles provided evidence for this chapter (**Figure 1**). Relevant findings from these fourteen articles have been included in the chapter and graded according to the GCDMP evidence grading criteria as described in **Table 3**. This synthesis of the literature relevant to the design and development of data collection instruments was performed to support transition of this chapter to an evidence-based guideline.

## Appendix A – CRF Printing

The following guidance is extracted and modified from the 2013 GCDMP chapter "CRF Printing and Vendor Selection" published in May 2007. Use of the following guidelines will help ensure the same quality and service from the contracted print vendor that the Clinical Data Manager expects to receive.

### CRF Binder

Prior to submitting the final printing specifications to the printer, the final print-ready CRF and shipping/distribution timetable should be approved by appropriate project team members. CRF binder specifications should include all of the information the vendor needs to produce the CRF binder and associated materials.

To determine the total number of CRFs, diaries or other required pages to be printed, consider the number of evaluable participants required per the protocol, the expected drop-out/replacement rate, and the possible need for a back-up supply. The back-up supply should be 10–15% of the total number of participants enrolled. If materials are distributed in packages, overage estimates should take into account the extra items that are in the pack. For example, if SAE forms are printed on a pad of 100 forms, they will be distributed in allotments of 100. Generally, a site that requires 101 pages will actually use 200 printed forms.

Also estimate the number of CRF pages with a breakdown of the number of no-carbon-required (NCR) pages, non-NCR pages, and other pages (e.g., diary or quality of life pages).

### Paper

Specify the paper to be used for printing the CRFs. Include information on the type of paper, color, page weight, hole-punch, perforation, and gum for each page or section. For example, conventional three-part, NCR paper comes in many colors and weights. The type and number of NCR pages required depend on the workflow and system used. Scanning or fax-based systems may require only two copies (the original white copy for scanning and the site copy).

There are other special considerations with the use of NCR paper. Printer specifications should include a piece of cardboard or other provision for the site to protect unused pages while completing a CRF page. When using a new vendor or a new paper supplier, it is advisable to test the NCR paper. The copy quality on the second or third ply is dependent on the quality of NCR paper. The weight of the paper should also be specified depending on your workflow. Paper of certain weights has been known to work more efficiently when faxed or scanned. If evaluating the paper supplied by a vendor, test the paper's quality when used to fax or scan printed material.

If adverse events and medications are collected at each visit and then extracted at every monitor visit, a pull-page system may be used. For example, a clinical data manager may use four-part NCR paper in which the fourth page is extracted first (a pull page), thereby enabling the data to be collected earlier. In an alternative approach, the fourth copy could be non-NCR so the next copy of the document reflects only the changes to the data.

### Tab Banks

Tab banks are very helpful to sites in navigating the CRF during the clinical study. Specify the number of tab banks and number of tabs per bank. Organizing the printing specifications by tabs can effectively communicate the collation order to the printer. Also, specify the paper weight of the tabs (usually card stock), the type and color of Mylar dip or other laminate on the tabs, and the text to be printed on each tab or tab page.

### Binding, Packaging, and Shipments

Specify the type of binding, binder color, width, number of inside pockets, cover text or art, and spine label. Specify the packaging instructions and include a packing list of the items that each site should receive. For example, special forms such as drug accountability logs, screening logs, SAE forms, diaries, and questionnaires may be bound separately in books or pads. Special forms may also be conveniently shrink-wrapped in appropriate numbers for each site.

If the printing vendor is shipping materials to sites, provide shipping instructions. Specify the number of sites and the number of items per site, the shipping company, and the shipping method (eg, ground or air). When finalizing timelines, the location of sites should be considered. Shipping to international sites may require additional time. With the shipping timetable, provide process instructions for tracking the shipment, checking the inventory of the shipment, and notifying the sponsor of the shipment's status.

### Information Commonly Provided With Printing Specifications

If applicable, the following information should be provided to the printer in addition to the printing specifications:

- The final camera-ready artwork of the CRF, the diary, and other pages in electronic files. The format of any electronic files should be discussed and agreed upon with the printing vendor.
- The specifications for CRF layout (e.g., identifying location of tabs, instructions on the back of tabs, collation of pages, etc.).
- A list of tabs, including the breakdown by bank and color.
- The camera-ready artwork of instructions to be printed on the tab backs.
- The company logo and text for the spine label.
- If the printing vendor is shipping to the sites, a list of sites and their mailing addresses. Moreover, shipping instructions should include details on how the printer will know when the site is approved to receive study materials.
- The priorities and specifications for printing the barcode, if applicable.
- The tentative timetable for sending the final master copy to the printer, for reviewing the materials prior to the final printing run, and the deadline for the arrival of the shipments at the sites.

The printer should provide a complete prototype of the CRF book for review and approval before the final print run. The prototype should include all of the book's pages and tabs, the spine label of the book, and the cover of the book.

New printing specifications (including printing and shipping timetables) should be submitted to the printers each time significant modifications are made to the CRF or to any item outlined in the specifications. An example of a CRF printing specifications checklist appears on the next page.

***Sample CRF Printing Specifications Checklist***

Total # of CRF binders to be printed _____

Total # of diaries to be printed _____

Total # of CRF pages per binder _____

# of NCR pages per binder _____

# of non-NCR pages per binder _____

# of diary pages per binder _____


Page formats: 2-part NCR with 2nd part cardstock, or specify other (The first part NCR should be white paper of weight 26):

_____


Specify page format for diary pages and diary covers (ex. Tri-fold):

_____

_____


Tabs: specify # of banks, # tabs/bank, #tabs with printed instructions on back, mylar-laminated or not and Mylar color:

_____


Does printer need to add page numbers?: Y N


Binders (specify):


Color: _____ Attach spine label

## Version History

| Date | Revision description |
| --- | --- |
| September 2000 | Initial publication. |
| May 2007 | Revised for style, grammar, and clarity. Substance of chapter content unchanged. |
| October 2010 | Revised for content, style, grammar, and clarity. Chapter title changed from "Data Acquisition" to "Design and Development of Data Collection Instruments." |
| TBD | Major revision to account for practice and regulatory changes |

## Competing Interests

The authors have no competing interests to declare.

## References

1. **Society for Clinical Data Management.** Good clinical data management practices (GCDMP). October 2013. Accessed November 3, 2025. https://scdm.org/wp-content/uploads/2024/04/Full-GCDMP-Oct-2013.pdf

2. **Journal of the Society for Clinical Data Management.** Good clinical data management practices (GCDMP). 2021. Accessed November 3, 2025. https://www.jscdm.org/collections/4/

3. **ICH Harmonised Guideline for Good Clinical Practice E6(R3).** International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Published January 6, 2025. Accessed November 3, 2025. https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf

4. **Medicines and Healthcare products Regulatory Agency (MHRA).** 'GXP' data integrity guidance and definitions, revision 1. March 2018. Accessed November 3, 2025. https://www.gov.uk/government/publications/guidance-on-gxp-data-integrity

5. **Food and Drug Administration, US Department of Health and Human Services.** General principles of software validation; guidance for industry and FDA staff, January 2002. Accessed November 3, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/general-principles-software-validation

6. **World Health Organization Expert Committee on Specifications for Pharmaceutical Preparations.** TRS 966 – Annex 5: WHO good data and record management practices, September 2016. Accessed November 3, 2025. https://www.gmp-compliance.org/files/guidemgr/WHO_TRS_996_annex05.pdf

7. **Food and Drug Administration, US Department of Health and Human Services.** Guidance for industry: providing regulatory submissions in electronic format – standardized study data, June 2021. Accessed November 3, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/providing-regulatory-submissions-electronic-format-standardized-study-data

8. **Food and Drug Administration, US Department of Health and Human Services.** Study data technical conformance guide – technical specifications document, June 2023. Updated March 2025. Accessed November 3, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/study-data-technical-conformance-guide-technical-specifications-document

9. **Pharmaceutical Inspection Convention, Pharmaceutical Inspection Co-operation Scheme (PIC/S).** PIC/S guidance good practices for data management and integrity in regulated GMP/GDP environments; PI 041-1, July 2021. Accessed November 3, 2025. https://picscheme.org/docview/4234

10. **European Commission Health and Consumers Directorate-General.** Health systems and products, Medicinal products – quality, safety and efficacy, Technical guidance on the format of the data fields of result-related information on clinical trials submitted in accordance with Article 57(2) of Regulation (EC) no 726/2004 and Article 41(2) of Regulation (EC) no 1901/2006, 2013. Accessed November 3, 2025. https://www.gmp-compliance.org/files/guidemgr/2013_01_22_tg_en.pdf

11. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), April 2016. Accessed November 3, 2025. https://eur-lex.europa.eu/eli/reg/2016/679/oj

12. **European Medicines Agency.** Guideline on computerised systems and electronic data in clinical trials EMA/INS/GCP/112288/2023. Accessed November 3, 2025. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-computerised-systems-and-electronic-data-clinical-trials_en.pdf

13. **Design and Development of Data Collection Instruments.** *Journal of the Society for Clinical Data Management.* 2023;1(1):22,1–7.

14. **CDISC (Clinical Data Interchange Standards Consortium).** Accessed November 11, 2023. https://www.cdisc.org

15. **Pomerantseva V.** Coding & designing a clinical database. *Appl Clin Trials.* 2012;21(3). Accessed November 3, 2025. https://www.appliedclinicaltrialsonline.com/view/coding-designing-clinical-database

16. **Is your eSystem actually an eCRF?** Accessed November 3, 2025. https://mhrainspectorate.blog.gov.uk/2021/05/11/is-your-esystem-actually-an-ecrf-electronic-case-report-form/

17. **Brembilla A, Martin B, Parmentier A,** et al. How to set up a database?—a five-step process. *J Thorac Dis.* 2018;10(Suppl 29):S3533-S3538. DOI: https://doi.org/10.21037/jtd.2018.09.138

18. **Bellary S, Krishnankutty B, Moodahadu L.** Basics of case report form designing in clinical research. *Perspect Clin Res.* 2014;5(4):159–166. DOI: https://doi.org/10.4103/2229-3485.140555

19. **Food and Drug Administration, US Department of Health and Human Services.** Data Integrity and Compliance with Drug CGMP, Questions and Answers Guidance for Industry (December 2018). Accessed November 3, 2025. https://www.fda.gov/media/119267/download

20. **Food and Drug Administration, US Department of Health and Human Services.** Conducting Clinical Trials With Decentralized Elements (September 2023) Accessed November 3, 2025. https://www.fda.gov/media/167696/download

21. **Farrow B.** In praise of skip logic, OpenClinica, 2019. Accessed November 3, 2025. https://www.openclinica.com/blog/in-praise-of-skip-logic/

22. **Jolley S.** Clinical Safety, Administration and Data Systems: How Should They be Integrated? *Drug Inf J.* 1995;29:661–663. DOI: https://doi.org/10.1177/009286159502900244

23. **Richesson R, Nadkarni P.** Data standards for clinical research data collection forms: current status and challenges. *J Am Med Inform Assoc.* 2011;18:341–346. Accessed November 3, 2025. DOI: https://doi.org/10.1136/amiajnl-2011-000107

24. **Fleischmann R, Decker A-M, Kraft A, Mai K, Schmidt S.** Mobile electronic versus paper case report forms in clinical trials: a randomized controlled Trial. *BMC Medical Research Methodology*, 2017;17:153. Accessed November 3, 2025. DOI: https://doi.org/10.1186/s12874-017-0429-y

25. **Food and Drug Administration, US Department of Health and Human Services.** Guidance for Industry: Computerized Systems Used in Clinical Investigations, May (2007). Accessed November 3, 2025. https://www.nus.edu.sg/research/docs/librariesprovider3/references/fda-computerized-systems-used-in-clinical-trials-2007.pdf?sfvrsn=b520a6c5_2

26. **Mohanty R, Gowda A, Nair A, Sharma A, Barick U.** Optimal eCRF design, user friendly interface and proper training: quintessential for high quality data in real world evidence (RWE) studies. *Int J Med Res Health Sci.* 2015;4(3):675. Accessed November 3, 2025. DOI: https://doi.org/10.5958/2319-5886.2015.00129.0

27. **Hufford M, Stokes T, Paty J.** Collecting reliable and valid real-time patient experience data. *Drug Inf J.* 2001;35(3):755–765. DOI: https://doi.org/10.1177/009286150103500314

28. **Meinecke A, Welsing P, Kafatos G,** et al. Series: Pragmatic trials and real world evidence: Paper 8. Data collection and management. *J Clin Epidemiol.* 2017;91:13–22. Accessed November 3, 2025. DOI: https://doi.org/10.1016/j.jclinepi.2017.07.003

29. **Kulis D, Bottomley A, Velikova G, Greimel E, Koller M.** EORTC Quality of life group translation procedure. European Organisation for the Research and Treatment of Cancer. Fourth Edition, 2017. Accessed November 3, 2025. https://www.eortc.org/app/uploads/sites/2/2018/02/translation_manual_2017.pdf

30. **CDISC.** COVID-19 Therapeutic Area User Guide for Version 2.0. Accessed November 3, 2025. https://www.cdisc.org/standards/therapeutic-areas/covid-19/covid-19-therapeutic-area-user-guide-v2-0

31. **Pétavy F, Seigneuret L, Hudson L,** et al. Global standardization of clinical research data. *Appl Clin Trials.* 2019;28(4). Accessed November 3, 2025. https://www.appliedclinicaltrialsonline.com/view/global-standardization-clinical-research-data

32. **Kubick W, Ruberg S, Helton E.** Toward a comprehensive CDISC submission data standard. *Ther Innov Regul Sci.* 2007;41:373–382. Accessed November 3, 2025. DOI: https://doi.org/10.1177/009286150704100311

33. **Nahm M, Shepherd J, Buzenberg A,** et al. Design and implementation of an institutional case report form library. *Clin Trials.* 2011;8(1):94–102. doi: Accessed November 3, 2025. DOI: https://doi.org/10.1177/1740774510391916

34. **Dugas M.** Design of case report forms based on a public metadata registry: re-use of data elements to improve compatibility of data. *Trials.* 2016;17(566). Accessed November 3, 2025. DOI: https://doi.org/10.1186/s13063-016-1691-8

35. **Food and Drug Administration, US Department of Health and Human Services.** Guidance for Industry, Electronic Source Data in Clinical Investigations (September 2013). Accessed November 3, 2025. https://www.fda.gov/media/85183/download

36. **CDISC eCRF Portal.** Accessed November 3, 2025. https://www.cdisc.org/kb/ecrf

37. **Food and Drug Administration, US Department of Health and Human Services.** Study Data Technical Conformance Guide, Guidance for Industry Providing Regulatory Submissions in Electronic Format – Standardized Study Data (Dec 2023). Updated March 2025. Accessed November 3, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/study-data-technical-conformance-guide-technical-specifications-document

38. **Cooper C, Buckman-Garner S, Slack M, Florian J, McCune S.** Developing standardized data: connecting the silos. *Drug Inf J.* 2012;46(5):521–522. Accessed November 3, 2025. DOI: https://doi.org/10.1177/0092861512454117

39. **International Bureau of Weights and Measures (December 2022), The International System of Units (SI) (PDF), vol.** 2 (9th ed.), ISBN 978-92-822-2272-0, archived from the original on 18

October 2021. Accessed November 3, 2025. https://www.bipm.org/documents/20126/41483022/SI-Brochure-9-EN.pdf

40. **Food and Drug Administration, US Department of Health and Human Services.** Investigational New Drug Application, Transfer of obligations to a contract research organization, 21 CFR §312.52 (1997). Accessed November 3, 2025. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-D/part-312/subpart-D/section-312.52

*Further Reading*

**Bassion S.** The Clinical Data Interchange Standards Consortium Laboratory Model: Standardizing Laboratory Data Interchange in Clinical Trials. *Drug Information Journal.* 2003;37(3):271–281. DOI: https://doi.org/10.1177/009286150303700303

**Chuang-Stein C.** Points for Consideration in the Collection and Analysis of Safety Data. *Drug Information Journal.* 1995;29(1):37–44. DOI: https://doi.org/10.1177/009286159502900106

**Hague D.** Changing platforms without stopping the train: experiences of data management and data management systems when adapting platform protocols by adding and closing comparisons. *Trials.* 2019;20(1):294. DOI: https://doi.org/10.1186/s13063-019-3322-7

**Lin CH.** A multi-technique approach to bridge electronic case report form design and data standard adoption. *J Biomed Inform.* 2015;53:49–57. DOI: https://doi.org/10.1016/j.jbi.2014.08.013

**Lin CH.** A standard-driven approach for electronic submission to pharmaceutical regulatory authorities. *J Biomed Inform.* 2018;79:60–70. DOI: https://doi.org/10.1016/j.jbi.2018.01.006

**Psotka MA.** Design of a "Lean" Case Report Form for Heart Failure Therapeutic Development. *JACC Heart Fail.* 2019;7(11):913–921. DOI: https://doi.org/10.1016/j.jchf.2019.07.001

**Richesson RL.** Use of SNOMED CT to represent clinical research data: a semantic characterization of data items on case report forms in vasculitis research. *J Am Med Inform Assoc.* 2006;13(5):536–46. DOI: https://doi.org/10.1197/jamia.M2093

**Somanath M.** Good practices in the conduct of a patient registry. *International Journal of Medical Research & Health Sciences.* 2016;5(4):22–26.

**Stout A.** NINDS Common Data Element Project – Approach and Methods. *Clin Trials.* 2012;9(3):322–329. DOI: https://doi.org/10.1177/1740774512438980