
Edit Check Design Principles

December 2009

Abstract

Edit checks are invaluable tools for increasing data quality and providing greater efficiency during data review and cleaning activities. This chapter discusses the process of edit check creation, including balance and efficiency considerations. The chapter also describes different types of edit checks, edit check validation, strategies for edit check specification creation, training related to edit checks, and considerations for using edit checks in studies that are paper based or use electronic data capture.

Introduction

The ultimate goal of clinical data management (CDM) is to complete every study with a dataset that accurately represents data captured in the study. No matter how much care is taken in collecting and entering data, discrepancies and data errors will invariably find their way into a clinical database. The vast majority of these data inconsistencies and errors can be alleviated with careful review and data-cleaning activities.

Review and cleaning of various data types may be performed by different personnel according to their knowledge and training. For example, data managers may not have the relevant medical knowledge to determine if an out-of-range lab value is indicative of a possible adverse event (AE) unless explicitly defined in the protocol or data management plan (DMP). Similarly, data entry personnel may not have the level of knowledge needed to recognize data indicative of protocol violations. Although responsibilities vary between organizations, CDM typically reviews triggered edit checks in addition to reviewing data that may be outside the scope of data entry personnel's experience.

Carefully designed edit checks can greatly increase efficiency and data quality by automating many data review processes within the clinical database or clinical data management system (CDMS). CDM personnel and members of the study team should collaborate to determine what edit checks should be in place to fulfill study requirements and reduce potential data errors and inconsistencies. Although assignment of responsibilities varies between organizations, CDM may be involved with all phases of edit check specification and testing, with the possible exception of edit check programming.

Scope

This chapter discusses the use of edit checks in clinical studies, including the purpose of edit checks, types of edit checks, creation processes of edit check specifications and development, and edit check testing. The chapter is intended as an overview of edit checks from a CDM perspective, and does not discuss details of programming and conditional statements used in edit checks.

Roles and responsibilities vary between organizations, and some of the topics discussed in this chapter may be the responsibility of different departments in different organizations. Regardless of role assignment, CDM personnel should be aware of the processes discussed in this chapter and how they impact their roles as data managers.

Minimum Standards

- Finalize protocol and complete initial database specifications prior to designing edit checks.
- Specify edit checks based on parameters of case report form (CRF) pages and safety and efficacy parameters from the protocol.
- Specify edit checks for all primary study endpoints and safety data.
- If applicable, specify edit checks with external data (e.g., laboratory data) for reconciliation purposes.
- Ensure all edit checks are programmed, validated, and documented in accordance with established standard operating procedures.

- Ensure all edit checks specification documents are appropriately version controlled.
- Provide training to relevant personnel on the impact of edit checks on their individual roles in entering and managing clinical data.

Best Practices

- Where appropriate, specify edit checks to compare study inclusion and exclusion criteria and any data (that are collected in CRF pages) that could be indicative of protocol violations.
- Design edit check specifications so redundant output does not occur when edit checks are run.
- Review edit checks with appropriate clinical and statistical personnel to ensure edit checks meet study needs and help identify inconsistencies in study endpoints.
- Specify edit checks for all study endpoints and all data supporting safety data and study endpoints.
- Develop a library of standard CRFs and edit checks based on standards used, such as CDASH or company-specific standards.
- Perform a quality control review of edit check design and specifications prior to performing user acceptance testing (UAT) of edit checks.
- Evaluate the effectiveness of edit checks once in active use, and modify, delete or create new edit checks accordingly.

Purpose and Process of Edit Checks

The purpose of edit checks is to draw attention to data that are inconsistent or potentially erroneous. Edit checks may be described as automatic warnings or notices that are generated by a database, CDMS, or other data entry application, and are triggered by data that are missing, out of range, unexpected, redundant, incompatible or otherwise discrepant with other data or study parameters. Most edit checks are triggered during the data entry

process, and may prompt the data entry operator to double check a value before saving the data. Other edit checks may be triggered by characteristics of related or aggregate data, and are more likely to notify CDM personnel of potential data errors after data entry has occurred. The potential data errors identified by triggered edit checks may prompt CDM personnel to perform data-cleaning activities such as performing self-evident corrections or generating queries to a site.

Balance and Efficiency Considerations

When creating edit check specifications, balance and efficiency considerations should be taken into account. Although edit checks can save considerable time and money in regard to data accuracy and cleaning, an edit check should not be created simply because it is possible to do so. Edit check specifications should be carefully designed to ensure checks are in place for critical data fields such as efficacy and safety variables. However, for variables not related to study endpoints or safety parameters, an evaluation should be made to determine whether the benefit provided by an edit check justifies the resources needed to create and test the edit check. This process should also evaluate the benefit of the edit check against the resources needed to review and close discrepancies generated by the edit check, as well as the resources needed to conduct the query process once the study is in progress.

When evaluating balance and efficiency factors for edit check specifications, consider that some edit checks may be less feasible or efficient than a manual review. For example, although edit checks can be created for free text data fields, manual review of listings by CDM may be more efficient, reliable, and cost-effective for this type of data. Even if an edit check could be programmed to account for every possible variant in a free text field (which is doubtful), a manual listing review would typically be more efficient and better suited to identifying unanticipated entries.

Some data irregularities may be more appropriately identified by biostatisticians than through edit checks or manual reviews. Some unexpected data trends may be indicative of systemic problems with data collection or processing and may not be easily identified by an edit check or manual review. In many cases, these types of data trends are most accurately and efficiently identified during preparation for statistical analysis. Because biostatisticians may also be able to suggest edit checks that can make their

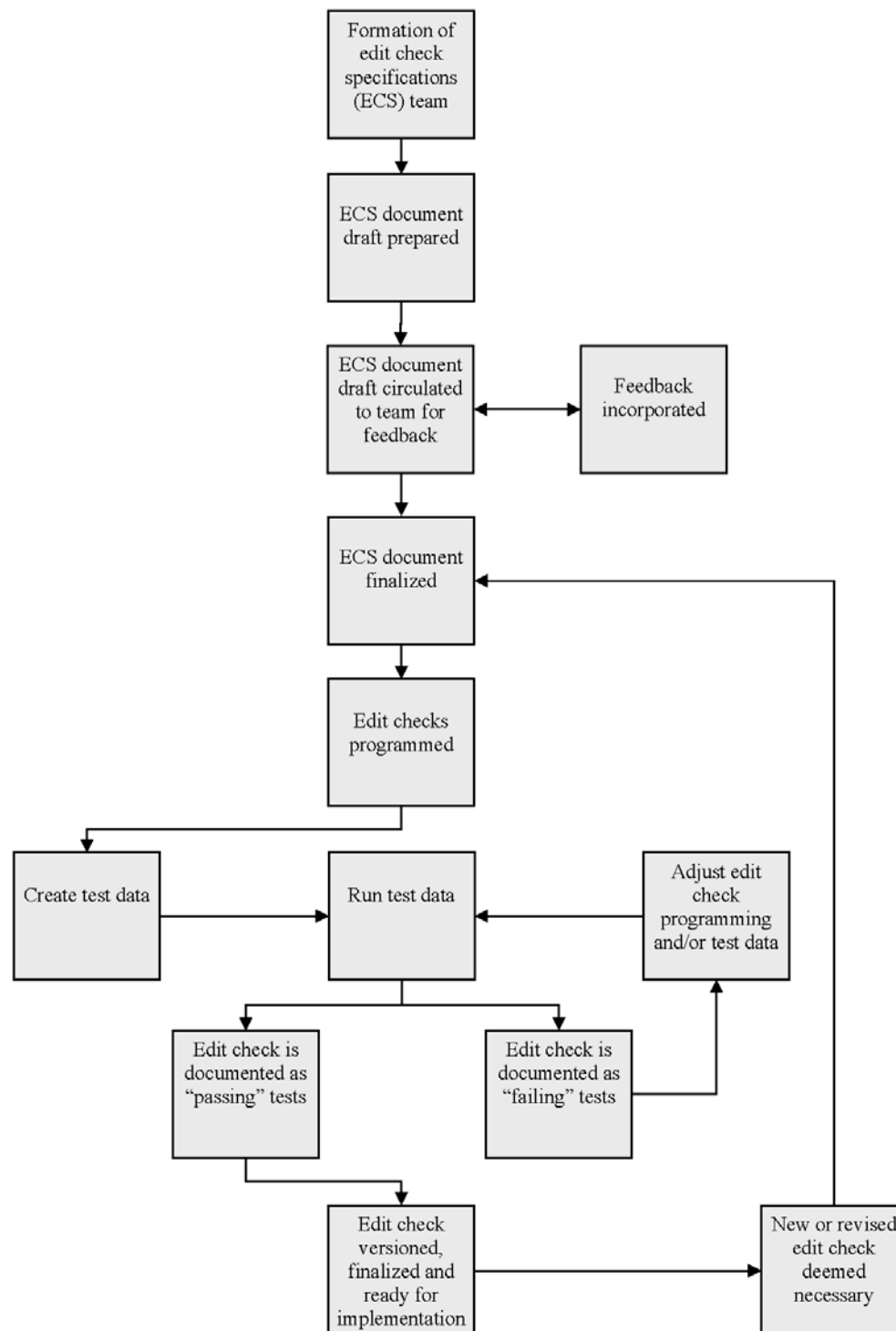
work more efficient when performing statistical analyses, consult the biostatistician(s) when designing edit check specifications.

Other potential data errors may be most efficiently identified by clinical research associates (CRAs), medical monitors, or medical coders. In many cases, a CRA or medical monitor may identify potential data errors by noting a trend and requesting a listing. Subsequent review of the requested listing may allow the CRA or medical monitor to confirm or deny the presence of the suspected data error(s). Additionally, medical coders may identify inconsistencies while coding data and subsequently bring these inconsistencies to the attention of appropriate CDM personnel.

Process of Edit Check Development

Edit check development is a process that requires information from a variety of sources, and should ideally incorporate a multidisciplinary approach to ensure appropriate and effective edit checks are implemented.¹ Although some details of edit check development processes may vary between organizations, the general steps should be similar between organizations for both edit check creation and testing. Figure 1 presents an overview of edit check development and testing processes.

Figure 1. Flowchart of Edit Check Development and Testing Processes



Creating Edit Check Specifications

Edit check specifications are crucial to identify invalid data, missing data, inconsistent data, and out-of-range values. Edit check specification planning requires information from a number of sources and should be performed with a comprehensive strategy for specification development in place prior to creating the initial draft.

Sources of information for edit check specifications may include:

- Study protocol—The study protocol describes the intent of a study, identifying inclusion/exclusion criteria, safety parameters, and primary and other study endpoints.
- Data management plan—Although the study protocol provides a broad overview of study parameters, a DMP typically describes in more detail data conventions for the study and identifies variables for which edit checks may need to be designed.
- Annotated CRFs and database design documentation—After identifying variables for which edit checks will be created, annotated CRFs and the database design should be examined to ensure edit checks are properly aligned with answer choices and the database structure.
- Standard edit check macros—Developing and maintaining a repository of commonly used edit check macros can save considerable time and money by avoiding duplication of work across studies or datasets.²
- Biostatisticians—Biostatistician(s) can provide direction regarding areas where edit checks may be desired to facilitate delivery of data that are suitable for statistical analyses without needing further cleaning or manipulation.
- Study personnel—Site personnel or other study personnel may be able to identify data fields that have been particularly prone to errors, inconsistencies, or out-of-range values in previous or similar studies.

Edit check specifications are typically documented in a table or spreadsheet format using various software applications. Although format, structure, and

level of detail may vary greatly between organizations, Table 1 presents an example of how an edit check specification table might be organized.

Table 1. Sample Edit Check Specification Table

CRF	Field Name (Number)	Check Name	Edit Check	Edit Check Message
ENROLL	Subject ID (2)	DUP_REC	Duplicate subject ID number	This subject ID number has already been assigned for this site. Please confirm correct ID number.
DEMOG	Subject ID (2)	NO_SUBJ_ID	Missing subject ID number	A subject ID number has not been entered for this record.
DEMOG	Subject DoB (6)	INVLD_AGE	Subject age is out of range	The date of birth value entered may be invalid. Please confirm correct date of birth.

Hierarchical View of Edit Checks

Some edit checks may be more important than others. Although a risk management approach can help identify edit checks that are crucial to the success of a study, a hierarchical approach to designing edit checks may be more efficient and provide similar results. The following items are an example of how a hierarchical sequence of edit check specification creation might be designed.

- General clinical data checks—These are checks designed to ensure key clinical data are accurate, reliable, and consistent. Although most edit checks fall under this category, some are more crucial than others.
 - Endpoint checks—Primary and other study endpoints should have checks in place to identify missing, erroneous, or out-of-range values. These are the variables for which statistical analyses will in part determine whether a study’s primary and secondary hypotheses are accepted or rejected. As such, the integrity of these data is crucial to the success of a study.
 - Safety checks—Edit checks should be created to help ensure any deviations from key safety parameters are noted and handled

accordingly. For example, if an AE is noted but no AE form is present, an edit check should flag this discrepancy so appropriate action can be taken.

- Protocol compliance checks—Data indicating adherence to study inclusion and exclusion criteria should be subjected to edit checks, as well as other protocol-specified parameters such as acceptable follow-up visit intervals.
- Programmed checks—For greatest efficiency, the majority of edit checks should be programmed into the clinical database or data capture system. These checks automatically trigger when certain predetermined conditions are met, such as missing data from a particular field or inconsistencies between data fields.
- Manual checks—Manual checks should be used for those data that cannot be easily checked through programmed edit checks, such as free text fields. Manual checks may also be used to verify key information such as site and subject identifiers on paper CRFs.
- Listings checks—Edit checks may also be designed for listings, which are used for checking multiple data points (where both correct and discrepant values may reside) across a single subject or module. Reviewing listings for discrepancies is typically a manual process.
- External checks—In some cases, most commonly in large complex studies, some checks may be programmed to run against data transferred from an outside source (e.g., labs). These checks are often run on multiple subjects with data from multiple datasets but only output data for subjects who fail the check.

Use of Standards for CRFs and Edit Checks

Use of standard edit checks based on standard CRF templates can save time and money while increasing quality, as well as potentially make the programming of edit checks easier. Use of standard edit checks can also decrease the amount of time needed for programming, therefore decreasing overall study timelines. Standard CRF templates may be prepared using CDASH or corporate standards, as appropriate, and version or change controls

should be applied. Although the types and scope of clinical studies may vary within a single organization, maintaining a central repository of CRFs and corresponding edit checks can reduce time and expenses for subsequent studies.

Standard edit checks should clearly identify the version of the corresponding standard CRF template. If a standard edit check template needs to be customized in some areas (e.g., a page number must be specified), the customized area should be flagged to draw attention to it. For more information about the CDASH standard, see the GCDMP chapter entitled “Data Management Standards in Clinical Research.”

Consistency in the Edit Check Specifications Document

The edit check specifications document should be consistent in its wording and conventions. The specifications document should also be consistent with the CRFs for which the edit checks are specified. The following are some examples of areas that should be reviewed for consistency within an edit checks specifications document.

- Use generic terms, such as "Subject" rather than "Patient," although a global change to “Patient” may need to be made for some studies.
- Note field names exactly as they are provided on the corresponding CRF (e.g., "Date of Birth" rather than "Birth Date," if “Date of Birth” is how the field is identified on the corresponding CRF).
- All descriptions in the edit check specifications document should be stated in complete sentences, using consistent terms such as "Visit Date must be present," or "If Not Done is marked, Result must be blank."
- Use consistent formatting conventions such as capitalizing all field names, or adding brackets only when a sentence is not clear without them (e.g., “A response must be marked for [Were any Adverse Events experienced?]”).
- Note any exceptions or special instructions for the reviewer (e.g., “NOTE: Do not query if page is lined through.”).

Message Wording

In addition to the care that must be taken to ensure edit checks are in place for key variables, the wording output by edit checks should be clear, unambiguous, and not leading. Any manually added queries to a clinical site should follow the same conventions as edit check output wording. The wording of both queries and edit check output messages should be carefully chosen to clearly and unambiguously relay the following information:

- Study, site, and subject or subject record—While adhering to data privacy conventions and regulations, queries and edit check outputs should clearly identify the study, site, and subject record for which an edit check or query is triggered.
- Variable name and value—Queries and edit check outputs should clearly identify what field, variable and value triggered the edit check or query.
- Supporting values—If an edit check or query is triggered from a derived value or is associated with other fields, supporting values should also be identified. For example, if an edit check is triggered by an out-of-range value for computed body mass index, the output message should indicate the value's relationship to the supporting fields containing subject height and weight.
- Message composition—Queries and edit check output messages should clearly identify the discrepant data and acceptable options for discrepancy resolution, but should not introduce bias or pose leading questions in any way. For example, an edit check for blood pressure should not output a message that specifies the expected range. Rather, the message should simply state that the value is out of the expected range and request confirmation or correction of the blood pressure.

Types of Checks

Edit checks are created to identify a number of different types of data inconsistencies or potential data errors. Although most edit checks are programmed into the database or CDMS and are triggered automatically when predefined conditions are met, data inconsistencies and potential data errors may also be found through manual data review.

Some of the most commonly used types of programmed edit checks include the following:

- Missing values—Edit checks for missing values are not usually applied to all data fields, but should be used for critical variables such as site and subject identification numbers or primary safety and efficacy variables.
- Missing CRF pages—In contrast to edit checks for missing values, edit checks for missing CRF pages may be applied to all CRFs. The intent of these checks is to highlight that an entire page or multiple pages have not been entered, which may be an oversight by the data entry operator or may result in a query to the site.
- Range checks—These are some of the more commonly used edit checks, and are intended to identify values that may be the result of an entry error or that may be indicative of a value outside of those expected for the subject population. Some examples may include height, weight, blood pressure, and other physiological parameters for which a particular range of values might be expected.
- Checks for duplicates—These checks are intended to negate the potential for the same data to be entered into the database more than once. Duplication may take the form of a duplicate subject identification number being used, a follow-up form being entered twice for a particular subject and interval, a single AE being entered twice, or any other situation where duplicate pages or data are entered.
- Logical inconsistencies across single CRF—The nature of potential logical inconsistencies may vary greatly between studies, but one example would be a CRF indicating that the subject is pregnant, but also indicating the subject is male. An edit check for this type of logical inconsistency can flag a data error that may not have been noticed otherwise.
- Inconsistencies across CRF pages or modules—Edit checks for logical inconsistencies are not limited to inconsistencies on a single CRF. Edit checks can also be programmed to identify discrepant data across CRF pages or modules. An example could be an edit check flagging an AE form that indicates that a medication was prescribed without the medication being recorded on a corresponding concomitant medications form.

- Checks of external data—Programmed edit checks are not limited to CRF data, but may also be applied to external data (lab data, ECG data, etc.). Many of these types of checks are primarily designed to help ensure that external data are consistent with the subject data within the database.
- Protocol violations—These checks are designed to identify specific data that may be indicative of protocol violations, and may take the form of range checks. One example would be calculating date ranges for follow-up visits to ensure all follow-up visits were within protocol-specified time windows. Another example would be checking subject eligibility forms to ensure all inclusion criteria were met and no exclusion criteria were met.

Front-End vs. Back-End Edit Checks

Edit checks that are triggered upon data entry are often referred to as front-end edit checks, whereas edit checks across multiple forms are often known as back-end edit checks. Front-end edit checks are typically limited to a single field or CRF page. An example of a front-end edit check would be a flag or warning that appears when an entry operator attempts to enter an impossible visit date, such as February 30 or a date in the future. Although front-end edit checks are usually more numerous, back-end edit checks are typically more complicated and therefore more difficult to program. An example of a back-end edit check would be one that notifies CDM personnel that a BMI (body mass index) entry is not consistent with the subject's reported height and weight.

Although details vary between studies and organizations, Table 2 presents which types of edit checks are more likely to be implemented as front-end checks, back-end checks, or both.

Table 2. Comparison of Edit Check Types

Type of check	Front-end check	Back-end check
Missing values	X	
Missing CRF pages	X	X
Range checks	X	
Checks for duplicates	X	X
Logical inconsistencies across single CRF	X	
Inconsistencies across CRF pages or modules		X
Checks of external data		X
Protocol violations	X	X

Electronic Data Capture (EDC) vs. Paper-based Edit Checks

Edit checks used in paper-based studies may differ somewhat from those used in EDC studies. For paper-based studies, some organizations may choose to limit the number of front-end checks. This ensures that potentially critical errors or discrepancies will be addressed directly by qualified CDM personnel. For studies using EDC, checks for transcription errors are not as necessary. However, more care must be taken in EDC studies to ensure the data entry design and front-end edit checks catch potential errors as they are entered. Because the electronic record may be considered the source document in some situations, there may be no other documentation to check against if possible errors are discovered later. The potential lack of additional source documentation in EDC studies also increases the importance of ensuring all edit checks are in place prior to the start of data collection. For more details about edit checks in studies using EDC, see the GCDMP chapter entitled “Electronic Data Capture—Concepts and Study Start-up.”

Validating Edit Checks

As with other aspects of a clinical database or EDC system, edit checks should be thoroughly tested and validated. Details may vary between different organizations and electronic systems, but the following process gives an overview of how edit checks should be validated.

- Creating test data—After edit checks are programmed, a set of test data should be created to mimic the type of data that are expected during the

study. This test data should not only include expected values, but also missing values and values that are out of range or that may not be expected from actual study data. These test data are typically created by CDM, although in some organizations database programmers or a quality assurance department may also be involved.

- Testing edit checks with test data—The test data used should include out-of-range or discrepant values that should trigger edit checks, as well as within-range or consistent values that should not trigger edit checks. The test data should contain all different scenarios that can occur for that check. For example, if an edit check is testing for a blood pressure range that is not between 80 mmHg and 200 mmHg, the out-of-range test data should ensure the edit check is triggered for anything below 80 and anything above 200 while the in-range data ensures the edit check is not triggered for anything between 80 and 200, including values that are exactly 80 or 200.
- Testing feedback loop process—This process may vary between organizations and is dependent upon who is doing the testing, which is usually the responsibility of CDM and programming personnel. CDM may give programming personnel the test plan and have the programming personnel test edit checks against the plan. If something in the plan does not occur as expected (e.g., an edit check was not triggered when it was supposed to or was triggered when it was not supposed to), the programmer notifies CDM, who may then modify the test data or add additional test data. Regardless of who performs the testing, data management should attempt to ensure all possible scenarios are tested, and should clearly document if any possible scenarios are not tested.
- Documentation—Every step of the edit check testing and validation process should be thoroughly documented. Both test data and edit checks may be documented on electronic or paper CRFs. If an organization does not have a formal test plan, these annotated test CRFs may suffice as a test plan provided the edit checks are described in sufficient detail. Documentation should also exist from the database showing where checks were triggered or not triggered. Any changes made to edit checks or test data during testing should also be documented. How documentation is achieved varies between organizations. One approach is to consolidate the test plan with the edit check specifications document, including a

“pass/fail” column that must be initialed and dated by the individual who is testing edit checks. Another approach may be to compile a binder with edit check specifications, the programming code behind edit checks, and test output from the database showing where each check was triggered or not triggered, with the initials and date of the individual who tested the check.

- **Quality control (QC)**—Although QC responsibilities may vary between organizations, some form of QC should be performed for the entire edit check validation process, final edit check programming, and all associated documentation. In different organizations, some or all of these QC processes may fall under the responsibilities of CDM personnel, project managers, database programmers, quality assurance personnel, or a manager of database development.
- **Validation of new or revised edit checks**—If any edit checks are added or revised during the course of a study, the same steps should be followed as are used for edit checks created at the beginning of the study.

Maintenance of Edit Checks

After edit check testing and validation has been completed, all responsible parties should provide written approval of edit check documentation prior to using the edit checks with actual subject data. CDM typically maintains an edit check document, ensuring that the document is kept current and incorporates proper version or change control. If substantial changes are made to the edit check document or the study is ongoing for more than a year, prior to study closeout CDM may request an additional review and approval of the final edit check document or changes made to the document. This re-review is intended to ensure that the needs of all parties continue to be met.

The edit check document should be considered a living document throughout the life cycle of the study. Edit checks may need to be changed as a result of CRF changes, or errors discovered in logic or terminology that need to be corrected. In addition, database programmers may suggest changes that result in more efficient data processing. As data are processed, new checks may be designed to identify discrepancies noted by monitors, biostatisticians, or other reviewers.

Change Control

Ideally, all changes to edit checks should be tracked within a single edit check document. However, a separate document may be employed if needed. All changes should be accompanied by the responsible individual's initials, the date of the change, and the reason for the change. If a change was approved or directed via e-mail, the date and sender of the e-mail should also be identified within the change document. Any new or changed edit checks should be thoroughly tested in accordance with established edit check testing procedures.

Version Control

Although different organizations may employ different strategies, a common strategy is for the first approved version of an edit check document to be considered Version 1 (V1.0). With this approach, minor administrative changes may be made at any time, and will change the version number by one-tenth (e.g., V1.1, V1.2, etc.). If CRF changes or other substantial changes occur, when the edit check document is subsequently updated, the version is updated by 1 (e.g., V2.0, V3.0). Regardless of the specific methodology used, all versions of an edit check document should be clearly documented.

Upon conclusion of a study, the final version of the edit check document should be archived with all other pertinent study documentation.

Edit Check Training

All data entry and CDM personnel who will be entering data, reviewing data, or reviewing the output of edit checks should be trained prior to data entry into the database. All personnel involved with these processes should have basic training in the formats, terminology, and use of edit checks, and the documentation of this training should reside in training folders. Training can be tailored to each individual role. For example, data entry personnel may only be trained on those edit checks that may be triggered upon data entry.

Data entry and CDM personnel may also need to undergo study-specific training for any edit checks that are unusual or unique to the study. If needed, a brief overview of the study and a review of the CRF may be included in the training. Study-specific training should also have clear documentation, and

may be maintained in training folders if confidentiality is not a concern. Otherwise, documentation of study-specific training may be maintained by data management and archived with all other pertinent study documentation at the close of the study.

Recommended Standard Operating Procedures

- Database Design
- Edit Check Specifications
- Edit Check Validation

References

1. Rondel RK, Varley SA, Webb CF, eds. *Clinical Data Management Second Edition*. West Sussex, England. John Wiley & Sons Ltd. 2000:113–119.
2. Gupta S. Standards for Clinical Data Quality and Compliance Checks. *Data Basics*. Winter 2008.

Further Reading

N/A

Chapter Revision History

Publication Date	Comments
December 2009	Initial publication.