German Medical Data Sciences 2024
R. Röhrig et al. (Eds.)

doi:10.3233/SHTI240843

Visualising Data Models of Patient Registries and Clinical Studies - A Method for Quality Check of EDC Systems

Beatrice COLDEWEY[a,1,] Philipp HONRATH[b], Stefan WOLKING[b],
Anna NIEMEYER[c], Rainer RÖHRIG[a,c], Yvonne WEBER[b],
and Myriam LIPPRANDT[a]
[a] *Institute of Medical Informatics, RWTH Aachen University, Germany*
[b] *Department of Neurology, Epilepsy Section, RWTH Aachen University, Germany*
[c] *TMF – Technology, Methods and Infrastructure for Networked Medical Research, Berlin, Germany*
ORCiD ID: Beatrice Coldewey https://orcid.org/0000-0001-9564-9467

Abstract. Introduction: The configuration of electronic data capture (EDC) systems has a relevant impact on data quality in studies and patient registries. The objective was to develop a method to visualise the configuration of an EDC system to check the completeness and correctness of the data definition and rules. **Methods:** Step 1: transformation of the EDC data model into a graphical model, step 2: Checking the completeness and consistency of the data model, step 3: correction of identified findings. This process model was evaluated on the patient registry EpiReg. **Results:** Using the graphical visualisation as a basis, 21 problems in the EDC configuration were identified, discussed with an interdisciplinary team, and corrected. **Conclusion:** The tested methodological approach enables an improvement in data quality by optimising the underlying EDC configuration.

Keywords. Data quality, patient registries, data model, electronic data capturing, electronic case report form

1. Introduction

In studies and registries, the type of data collection has a relevant impact on the data quality and thus also on the quality of the studies or patient registries [1–5]. Data definition is a structured process based on the study or registry protocol. The data should be described in a catalog of items as part of the data management plan (DMP) [1,2].

The use of electronic data capture (EDC) systems, such as REDCap (Vanderbild University, [6]) or LibreClinica (ReliaTec GmbH, [7]), enables researchers to collect and manage data effectively with higher data quality than paper-based formats [4,8–10]. The configuration of electronic case report forms (eCRFs) goes beyond defining the data to be collected and the type of response, e.g., checkboxes or free text. The systems provide

[1] Corresponding Author: Beatrice Coldewey, Institute of Medical Informatics, University Hospital RWTH Aachen, Pauwelsstraße 30, D 52074 Aachen. bcoldewey@ukaachen.de

options for implementing filter questions, help texts, data validation rules, notes on missing entries and autocompletion [6,7,11].

On the one hand, these features increase efficiency and ensure data quality already during data entry [3,12]. On the other hand, unclear and ambiguous definitions, configuration errors or a poor layout of the EDC systems can lead to questions or answer options not being displayed by mistake or answers to questions not being entered [1,5,13]. To prevent incomplete or inaccurate data, such errors must be avoided.

Registries are often created by clinicians with limited methodological experience [2]. Even with extensive experience, errors can occur when setting up or adjusting the configuration. In this case, it can be important to obtain an overview of the parameters, formulas and plausibility checks stored in the EDC system as part of the quality assurance, validation of the EDC system configuration or troubleshooting. The display options for the configurations are limited. Graphical modelling can be advantageous compared to tabular overviews, especially when there is a large amount of data and dependencies to be tracked. Abstract information translated into visual representations supports perception, understanding and communication e.g. how the values relate to one other [14,15].

The objective of this work was to develop a method to easily and safely visualise the configuration of an EDC system to check the completeness and correctness of the data definition and rules.

2. Methods

2.1. Approach

The approach developed in this work comprises a multi-step methodological procedure (see **Figure 1**). **Step one:** Based on existing registries and studies, the variables defined in the EDC system, their properties and interdependencies are transferred to a graphical visualisation using directed graphs with nodes and labelled edges. The graphical representation provides a quick overview and is the basis for discussion of the data to be recorded. **Step two:** The completeness and correctness of the data and rules configured in the EDC system are evaluated based on defined criteria. Ambiguities are discussed and resolved in an interdisciplinary team. **Step three:** Identified problems are resolved by adjusting the EDC configuration and, if necessary, the registry protocol, DMP and catalog of items.

2.2. Evaluation of the approach using the EpiReg-registry

As proof of concept, the suitability and usability of the methodological approach have been tested on a registry for patients with genetic epilepsy (EpiReg). The EpiReg-registry is part of the Treat-ION research project (01GM2210B) [16]. The catalog of items comprises a minimal dataset and an optional dataset [17]. The EDC system REDCap is used for data collection.

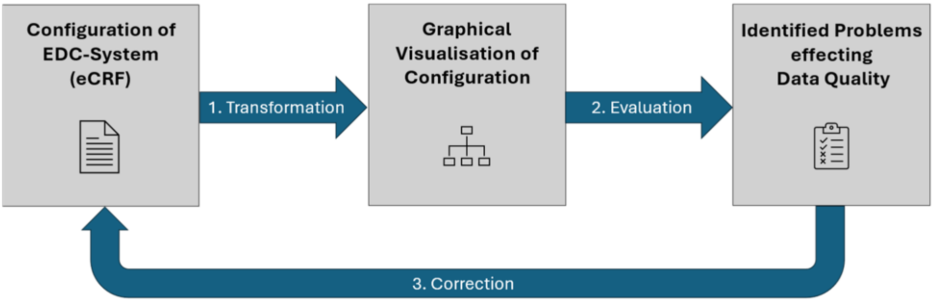

Configuration of EDC-System (eCRF)
1. Transformation
Graphical Visualisation of Configuration
2. Evaluation
Identified Problems effecting Data Quality
3. Correction

Figure 1: Methodical approach to optimise existing registries or studies by checking the completeness and correctness of the data definition and rules using a graphical visualisation

2.2.1. *Step one: Transformation to graphic visualisation*

Requirements have been specified with regard to the information to be displayed about the data elements in the EDC configuration. **Table 1** provides an overview of the necessary information and selected graphical modelling. The syntax is based on elements of directed graph modelling in software development, such as uml diagrams [18,19]. A simple visualisation tool (PowerPoint) was used during the initial testing. The implementation of a tool to automate the creation of the graphical visualisation was not carried out.

2.2.2. *Step two: Evaluation Criteria to improve data quality*

To counteract inadequate data quality due to incomplete or inaccurate data, a list of possible causes with regard to the configuration of the EDC system was generated (**Table 2**). The list included problems reported in literature [1,3–5,13] and was expanded by problems identified during the evaluation of the EpiReg registry. The assessment of the relevance for data quality (categorisation: ”low“, ”medium“, ”high“) is based on the risk associated with the error. The EpiReg registry was reviewed by a specialist in medical informatics. Identified problems were added as comments in the graphic and discussed in a joint meeting with two genetic epilepsy experts, two specialists in medical informatics and the person responsible for implementing the eCRF in REDCap.

2.2.3. *Step three: Correction of identified problems*

Identified problems are corrected in the configuration of the EDC system and published in a new version of the eCRF. If the identified problems were related to the underlying catalog of items, the data management plan was also adjusted.

Table 1: Description of the graphic elements and underlying requirements for the visualisation of the EDC configuration

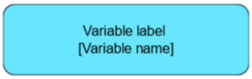
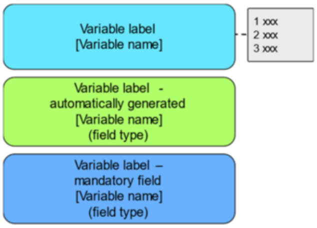
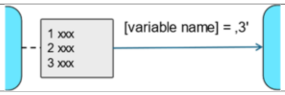
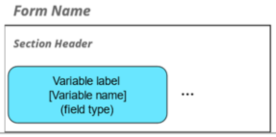
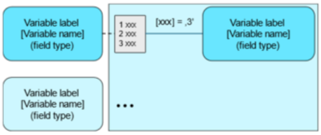
Requirement: The user must recognize ...	**Graphic Feature**	**Example graphic Visualisation**
which variables are being collected: Variable name (identifier) Variable label (short title)	Labelled node with variable label and variable name.	
which attributes the variables have: field type used? data validation rules? available choices? mandatory fields or automatically generated?	field type / acceptable values: additional information in the node with variable label choices: separate node connected via unlabelled edge. mandatory fields and automatically generated data: different colour coding.	 Variable label [Variable name] 1 xxx 2 xxx 3 xxx Variable label - automatically generated [Variable name] (field type) Variable label – mandatory field [Variable name] (field type)
how the variables are linked via the branching logic.	nodes connected via labelled edge	
how the variables are structured (form name and section header).	display of the form name as well as section header and framing of the associated variables	 Form Name Section Header Variable label [Variable name] (field type) ...
which variables are part of repeating structures.	colour coding of repetitive areas and affected variables, reduction to one-time display	 Variable label [Variable name] (field type) 1 xxx 2 xxx 3 xxx [xxx] = ‚3' Variable label [Variable name] (field type) Variable label [Variable name] (field type) ...

Table 2: Set of error categories and derived inspection criteria

Error category	**Inspection criteria**	**Visual instance**
Unclear / ambiguous data definitions [1]	Clear and concise questions, prompts, and instructions (suitable for target audience) [4,5] Unit of measurement /data format defined [4] Use of help texts and sample data [3,13] Choice "unknown" if data is not available	Nodes
Data overload [1]	No unnecessary data is collected	Nodes
Programming errors [1]	No missing or duplicate data elements /choices No error in branching logic (missing references / reference to incorrect variable/choice)	Nodes Edges Node tree
Poor CRF layout [1]	Structure - Structuring of the data (subdivision of questionnaires, subheadings, item order)[5,13] - Use of conditional questions (Hiding of questions) [3,13] Presentation - Selection of suitable field types (use precoded answer sets, minimize free text responses) [4,13] - Use of consistent formats, font style and font sizes, language [4,5,13]	Nodes Edges Section structure
Insufficient data checks [1]	Definition/consistent use of required items [3] Implementation of meaningful validation rules [3,5]	Nodes

3. Results

3.1. Graphical Visualisation

The data dictionary of the minimal dataset and the optional dataset of the EpiReg exported from REDCap were converted into a graphical visualisation using the defined visual elements. **Figure 3** shows the graphical model of the revised minimal dataset. The dataset contains 95 items, 14 of these are at the first level and 81 are dependent on choices of prior questions and linked to them via the branching logic. The request for information on up to five relevant genetic variants comprises 11 repeating questions and therefore 44 repetitive items that are not shown. The only automatically generated variable is the REDCap-ID. Three questions were marked as mandatory.

3.2. Examination and correction of the dataset

For the minimum dataset the version dated 21.09.2023 with 95 items was used as the basis for the review. The first evaluation and discussion of the data dictionary led to 11 documented problems (Table 3). **Figure 2** shows an excerpt of the adjustments made in comparison to the reviewed version of the minimal dataset. The problems were implemented in the revised version dated 02.01.2024 together with the extension and restructuring of some items (**Figure 3**). The review of the optional dataset with 988 items distributed over 11 questionnaires (version dated 23.02.2023) resulted in 10 problems which were included in a major revision of the dataset.

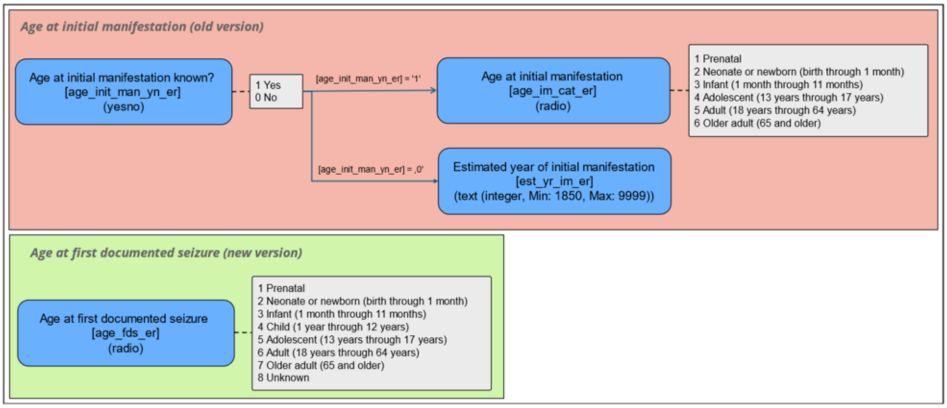
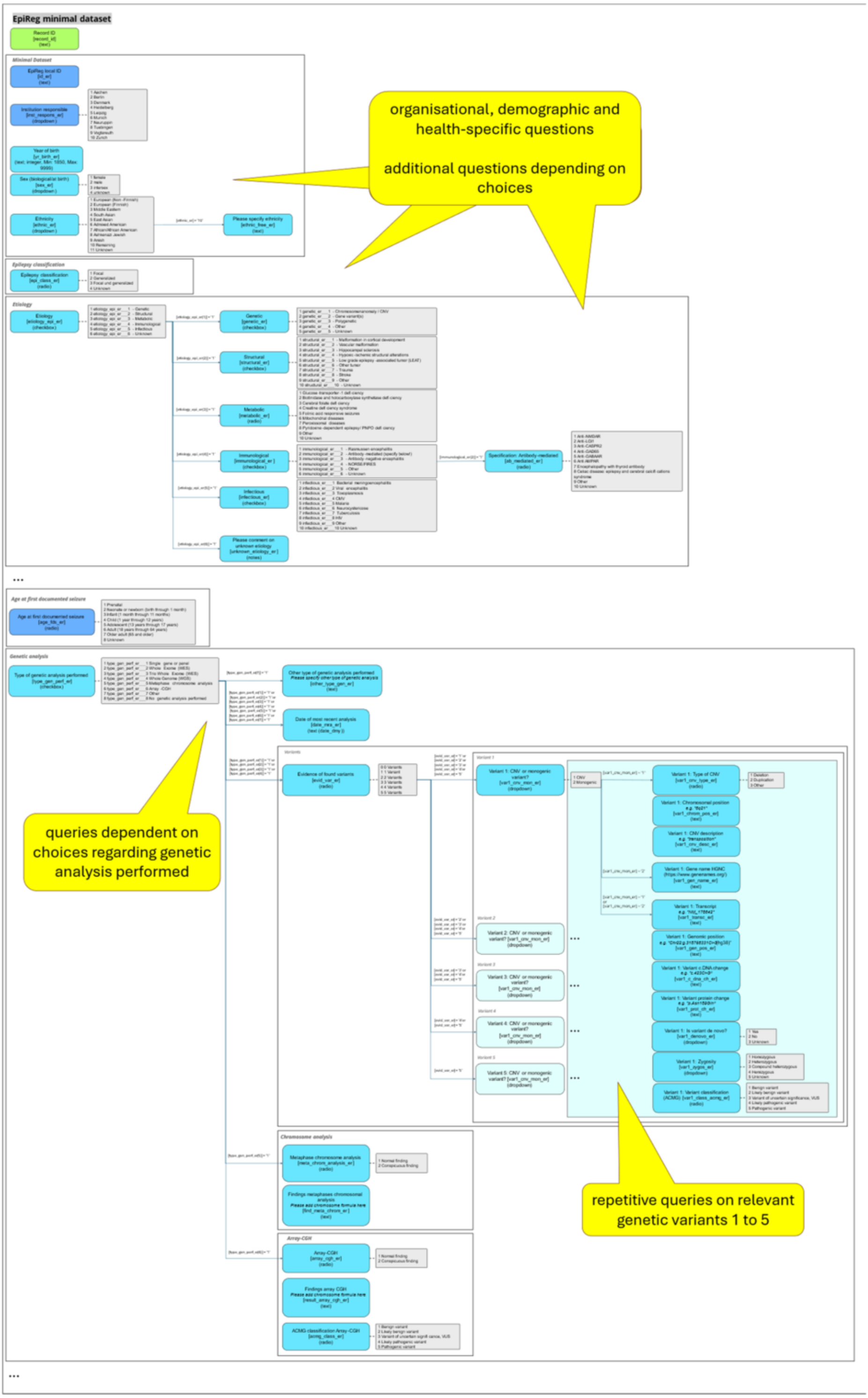

Age at initial manifestation (old version)
Age at initial manifestation known?
[age_init_man_yn_er]
(yesno)
1 Yes
0 No
[age_init_man_yn_er] = '1'
[age_init_man_yn_er] = ,0'
Age at initial manifestation
[age_im_cat_er]
(radio)
1 Prenatal
2 Neonate or newborn (birth through 1 month)
3 Infant (1 month through 11 months)
4 Adolescent (13 years through 17 years)
5 Adult (18 years through 64 years)
6 Older adult (65 and older)
Estimated year of initial manifestation
[est_yr_im_er]
(text (integer, Min: 1850, Max: 9999))
Age at first documented seizure (new version)
Age at first documented seizure
[age_fds_er]
(radio)
1 Prenatal
2 Neonate or newborn (birth through 1 month)
3 Infant (1 month through 11 months)
4 Child (1 year through 12 years)
5 Adolescent (13 years through 17 years)
6 Adult (18 years through 64 years)
7 Older adult (65 and older)
8 Unknown

Figure 2: Adjustments of question structure and extension of choices made to the "Age at first documented seizure" section (previously "Age at initial manifestation") to correct identified problems 3, 5 and 6 (see **Table 3**). As part of the revision, a renaming from "initial manifestation" to "first documented seizure" was carried out.

Table 3: Problems identified by analysing the EpiReg minimal dataset

	Error category	**Issue description**	**Risk**	**Corrective action**
1	Data Overload	Query of the date of the last change to the questionnaire: Adjustment can easily be forgotten and should therefore be automated or derived from the system-side change tracking	medium	Remove / change variable
2	Unclear/ambiguous data definitions,	Variable label “Unknown – please clarify“ for “Etiology”-branch not very self-descriptive: reference to parent item: “Please comment on unknown etiology”	low	Adjust label
3	Unclear/ambiguous data definitions, Poor CRF layout	Choice “unknown” should be explicitly selectable so that it can be distinguished from missing entries in the evaluation (n=12)	medium	Adjust choices
4	Programming errors	Branching logic for “Specification: Antibody-mediated” at 3rd level. Check can be shortened as it contains unnecessary queries (part of the check will always be true)	low	Adjust branching logic
5	Programming errors, Unclear/ambiguous data definitions	Missing choice for age group “Child (1 year through 12 years)” for item “Age at initial manifestation”	high	Adjust choices
6	Data Overload, Unclear/ambiguous data definitions	Conflicting query: “Estimated year of initial manifestation” after choice: “Age at initial manifestation known?"= “No”: No correct answer can be given	high	Adjust/remove variable
7	Poor CRF layout	Missing heading/bad position of headline for the subdivision of examination methods “chromosome analysis” and "Array-CGH"	low	Move/additional subheading
8	Programming errors, Unclear / ambiguous data definitions	Query of examination results after indication “not performed” of examination methods “chromosome analysis” and “Array-CGH”: No answer can be given	medium	Adjust choices and associated branching logic
9	Poor CRF layout	Note text “(multiple answers possible!)“ for item “Type of sequencing performed” has a different font size/type than for the other items	low	Adjust font size/type
10	Unclear / ambiguous data definitions	Choices for “Evidence of found variants” limited from 1 to 5 variants: option for “0 Variants” is missing	high	Adjust choices
11	Insufficient data checks	Change between required and not required items without recognisable reasons e.g. between different variables and same item	medium	Adjust required variables


EpiReg minimal dataset
Record ID [record_id] (text)
Minimal Dataset
EpiReg local ID [id_er] (text)
Institution responsible [inst_respons_er] (dropdown)
1 Aachen
2 Berlin
3 Denmark
4 Heidelberg
5 Leipzig
6 Munich
7 Neuruppin
8 Tuebingen
9 Vogtareuth
10 Zurich
Year of birth [yr_birth_er] (text, integer, Min: 1850, Max: 9999)
Sex (biological/at birth) [sex_er] (dropdown)
1 female
2 male
3 intersex
4 unknown
Ethnicity [ethnic_er] (dropdown)
1 European (Non-Finnish)
2 European (Finnish)
3 Middle Eastern
4 South Asian
5 East Asian
6 Admixed American
7 African/African American
8 Ashkenazi Jewish
9 Amish
10 Remaining
11 Unknown
[ethnic_er] = '10'
Please specify ethnicity [ethnic_free_er] (text)
Epilepsy classification
Epilepsy classification [epi_class_er] (radio)
1 Focal
2 Generalized
3 Focal und generalized
4 Unknown
Etiology
Etiology [etiology_epi_er] (checkbox)
1 etiology_epi_er___1 - Genetic
2 etiology_epi_er___2 - Structural
3 etiology_epi_er___3 - Metabolic
4 etiology_epi_er___4 - Immunological
5 etiology_epi_er___5 - Infectious
6 etiology_epi_er___6 - Unknown
Genetic [genetic_er] (checkbox)
1 genetic_er___1 - Chromosomenanomaly / CNV
2 genetic_er___2 - Gene variant(s)
3 genetic_er___3 - Polygenetic
4 genetic_er___4 - Other
5 genetic_er___5 - Unknown
Structural [structural_er] (checkbox)
1 structural_er___1 - Malformation in cortical development
2 structural_er___2 - Vascular malformation
3 structural_er___3 - Hippocampal sclerosis
4 structural_er___4 - Hypoxic-ischemic structural alterations
5 structural_er___5 - Low grade epilepsy-associated tumor (LEAT)
6 structural_er___6 - Other tumor
7 structural_er___7 - Trauma
8 structural_er___8 - Stroke
9 structural_er___9 - Other
10 structural_er___10 - Unknown
Metabolic [metabolic_er] (radio)
1 Glucose transporter-1 deficiency
2 Biotinidase and holocarboxylase synthetase deficiency
3 Cerebral folate deficiency
4 Creatine deficiency syndrome
5 Folinic acid responsive seizures
6 Mitochondrial diseases
7 Peroxisomal diseases
8 Pyridoxine-dependent epilepsy/ PNPO deficiency
9 Other
10 Unknown
Immunological [immunological_er] (checkbox)
1 immunological_er___1 - Rasmussen encephalitis
2 immunological_er___2 - Antibody-mediated (specify below!)
3 immunological_er___3 - Antibody-negative encephalitis
4 immunological_er___4 - NORSE/FIRES
5 immunological_er___5 - Other
6 immunological_er___6 - Unknown
Specification: Antibody-mediated [ab_mediated_er] (radio)
1 Anti-NMDAR
2 Anti-LGI1
3 Anti-CASPR2
4 Anti-GAD65
5 Anti-GABAAR
6 Anti-AMPAR
7 Encephalopathy with thyroid antibody
8 Celiac disease: epilepsy and cerebral calcifications syndrome
9 Other
10 Unknown
Infectious [infectious_er] (checkbox)
1 infectious_er___1 Bacterial meningoencephalitis
2 infectious_er___2 Viral encephalitis
3 infectious_er___3 Toxoplasmosis
4 infectious_er___4 CMV
5 infectious_er___5 Malaria
6 infectious_er___6 Neurocysticercose
7 infectious_er___7 Tuberculosis
8 infectious_er___8 HIV
9 infectious_er___9 Other
10 infectious_er___10 Unknown
Please comment on unknown etiology [unknown_etiology_er] (notes)
...
Age at first documented seizure
Age at first documented seizure [age_fds_er] (radio)
1 Prenatal
2 Neonate or newborn (birth through 1 month)
3 Infant (1 month through 11 months)
4 Child (1 year through 12 years)
5 Adolescent (13 years through 17 years)
6 Adult (18 years through 64 years)
7 Older adult (65 and older)
8 Unknown
Genetic analysis
Type of genetic analysis performed [type_gen_perf_er] (checkbox)
1 type_gen_perf_er___1 Single gene or panel
2 type_gen_perf_er___2 Whole Exome (WES)
3 type_gen_perf_er___3 Trio Whole Exome (WES)
4 type_gen_perf_er___4 Whole Genome (WGS)
5 type_gen_perf_er___5 Metaphase chromosome analysis
6 type_gen_perf_er___6 Array-CGH
7 type_gen_perf_er___7 Other
8 type_gen_perf_er___8 No genetic analysis performed
Other type of genetic analysis performed
Please specify other type of genetic analysis
[other_type_gen_er] (text)
Date of most recent analysis [date_mra_er] (text (date_dmy))
Variants
Evidence of found variants [evid_var_er] (radio)
0 0 Variants
1 1 Variant
2 2 Variants
3 3 Variants
4 4 Variants
5 5 Variants
Variant 1
Variant 1: CNV or monogenic variant? [var1_cnv_mon_er] (dropdown)
1 CNV
2 Monogenic
[var1_cnv_mon_er] = '1'
[var1_cnv_mon_er] = '2'
Variant 1: Type of CNV [var1_cnv_type_er] (radio)
1 Deletion
2 Duplication
3 Other
Variant 1: Chromosomal position e.g. "8q21" [var1_chrom_pos_er] (text)
Variant 1: CNV description e.g. "translocation" [var1_cnv_desc_er] (text)
Variant 1: Gene name HGNC (https://www.genenames.org/) [var1_gen_name_er] (text)
Variant 1: Transcript e.g. "NM_170842" [var1_transc_er] (text)
Variant 1: Genomic position [var1_gen_pos_er] (text)
Variant 1: Variant c.DNA change e.g. "c.423C>G" [var1_c_dna_ch_er] (text)
Variant 1: Variant protein change [var1_prot_ch_er] (text)
Variant 1: Is variant de novo? [var1_denovo_er] (dropdown)
1 Yes
2 No
3 Unknown
Variant 1: Zygosity [var1_zygos_er] (dropdown)
1 Homozygous
2 Heterozygous
3 Compound heterozygous
4 Hemizygous
5 Unknown
Variant 1: Variant classification (ACMG) [var1_class_acmg_er] (radio)
1 Benign variant
2 Likely benign variant
3 Variant of uncertain significance, VUS
4 Likely pathogenic variant
5 Pathogenic variant
Variant 2
Variant 2: CNV or monogenic variant? [var1_cnv_mon_er] (dropdown)
Variant 3
Variant 3: CNV or monogenic variant? [var1_cnv_mon_er] (dropdown)
Variant 4
Variant 4: CNV or monogenic variant? [var1_cnv_mon_er] (dropdown)
Variant 5
Variant 5: CNV or monogenic variant? [var1_cnv_mon_er] (dropdown)
...
Chromosome analysis
Metaphase chromosome analysis [meta_chrom_analysis_er] (radio)
1 Normal finding
2 Conspicuous finding
Findings metaphases chromosomal analysis
Please add chromosome formula here
[find_meta_chrom_er] (text)
Array-CGH
Array-CGH [array_cgh_er] (radio)
1 Normal finding
2 Conspicuous finding
Findings array CGH
Please add chromosome formula here
[result_array_cgh_er] (text)
ACMG classification Array-CGH [acmg_class_er] (radio)
1 Benign variant
2 Likely benign variant
3 Variant of uncertain significance, VUS
4 Likely pathogenic variant
5 Pathogenic variant
...

organisational, demographic and health-specific questions

additional questions depending on choices

queries dependent on choices regarding genetic analysis performed

repetitive queries on relevant genetic variants 1 to 5

Figure 3: Excerpt from the model of the minimum dataset of the EpiReg registry (Version 02.01.2024).

4. Discussion

The methodological approach aims to improve data quality in studies and registries at the time of entry by avoiding e.g. programming errors or poor CRF layout. Graphical modelling of the data dictionary provides a clear representation of all variables, their attributes and links between them and serves as the basis for the revision. The established list of test criteria enables the examination of quality and inconsistencies regardless of the underlying medical domain.

The evaluation of the approach using the EpiReg-registry for patients with genetic epilepsy as proof of concept enabled the identification of several issues. Correcting the identified issues can contribute the data quality and thus improve the interpretability of the registry data, which in many cases can have a significant impact on medical care.

Compared to a standard structured expert review of the eCRF using the provided abstract information, the graphical visualisation provided a more efficient way to perceive the information [14,15] and made it easier to check the quality criteria and to identify errors. The tree structure made it particularly easy to identify erroneous links or deviations in repetitive areas. In further studies, the superiority with and without graphical visualization is to be tested in comparison.

The graphical representation also simplified the discussion within the interdisciplinary team. Using the graphical model made it possible to review and discuss the data from both a medical and technical perspective. For example, not solely problems like programming errors in the branching logic were identified. The use of the alternative visualisation approach also led to an adjustment of e.g., the inclusion and sequence of individual data elements for reasons of medical relevance. To additionally extend the usefulness of the visualisation, the presentation could be supplemented by changes to the data set for better traceability.

The approach complements the quality assurance process of the multitude of evaluation/quality criteria that lead to high-quality registries and studies [1,2,11]. To ensure a comprehensive list of errors, the data should be checked by two independent reviewers and their results combined. A new check should be carried out after each revision. The basis for implementation in the EDC systems should always be a prior definition of the items to be collected. For a comprehensive optimisation of data dictionaries, the approach should be complemented by further methods, e.g. pre-tests with persons responsible for data entry, as part of the structured process.

The methodological approach has so far only been tested on one registry as an example. To further verify the generalisability of the method, the approach should be applied to the data dictionary of other registries and studies implemented in different EDC-Systems. The list of test criteria and graphical elements needed does not claim to be exhaustive. Further test criteria and elements for a comprehensive model should be added.

The graphical modelling has been performed manually using simple tools as part of initial tests. Since this is not feasible, the next step is to develop a tool that enables the automatic creation of graphical model, taking into account the functional scope of different EDC systems.

5. Conclusion

The chosen methodological approach provides criteria for checking and improving the eCRF of registries or studies based on a clear, graphical visualisation. In an exemplary use case, the approach has already contributed to the improvement of the eCRF of a registry and thus to data quality. Further development of a tool to automatically generate graphic visualisation and the addition of further test criteria are desirable.

Declarations

Ethical Vote: Not applicable for this work, Ethical Vote for EpiReg: Ethics Committee of RWTH Aachen University, Vote-No: EK172/21, Date: 17.6.2021 Chair: Prof. Dr. med. G. Schmalzing

Funding: BMBF funded project Treat-ION (01GM2210B) and DFG funded Research Unit FOR2715 WE4896/4-2

Conflict of Interest: All authors state that they have no conflict of interest.

Contributions of the authors: Conceptualization, Methodology: BC, ML, Investigation: BC, PH, SW, YW, ML, Data Curation: BC, ML, Resources: ML, YW, RR, Visualisation: BC, ML, Supervision: AN, RR, ML, Writing the initial draft: BC, Writing-Review & Editing: all authors.

References

[1] Arts DGT, Keizer NF de, Scheffer G-J. Defining and Improving Data Quality in Medical Registries: A Literature Review, Case Study, and Generic Framework. J Am Med Inform Assoc. 2002;9:600–611.

[2] Gutachten zur Weiterentwicklung medizinischer Register zur Verbesserung der Dateneinspeisung und -anschlussfähigkeit. [place unknown]: Bundesministerium für Gesundheit; 2021.

[3] Hoeijmakers F, Beck N, Wouters, Michel W J M, et al. National quality registries: how to improve the quality of data? J Thorac Dis. 2018;10:S3490-S3499.

[4] Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. Perspect Clin Res. 2014;5:159–166.

[5] Johnson CM, Nahm M, Shaw RJ, et al. Can prospective usability evaluation predict data errors? AMIA Annu Symp Proc. 2010;2010:346–350.

[6] Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42:377–381.

[7] ReliaTec GmbH. LibreClinica [Internet] [cited 2024 Apr 22]. Available from: https://www.libreclinica.org/index.html.

[8] Zeleke AA, Naziyok T, Fritz F, et al. Data Quality and Cost-effectiveness Analyses of Electronic and Paper-Based Interviewer-Administered Public Health Surveys: Systematic Review. J Med Internet Res. 2021;23:e21382.

[9] Le Jeannic A, Quelen C, Alberti C, et al. Comparison of two data collection processes in clinical studies: electronic and paper case report forms. BMC Med Res Methodol. 2014;14:7.

[10] Rorie DA, Flynn RWV, Grieve K, et al. Electronic case report forms and electronic data capture within clinical trials and pharmacoepidemiology. Br J Clin Pharmacol. 2017;83:1880–1895.

[11] Lindoerfer D, Mansmann U. A Comprehensive Assessment Tool for Patient Registry Software Systems: The CIPROS Checklist. Methods Inf Med. 2015;54:447–454.

[12] Naziyok TP, Feeken C, Zeleke AA, et al. Data Collection of Medication - Impact of Autocompletion in eCRFs on Efficiency and Data Quality. Stud Health Technol Inform. 2017;243:70–74.

[13] Minto C, Vriz GB, Martinato M, et al. Electronic Questionnaires Design and Implementation. Open Nurs J. 2017;11:157–202.

[14] Krause A, OConnell M. A Picture is Worth a Thousand Tables: Graphics in Life Sciences. [place unknown]: Springer US; 2016.

[15] Cleveland WS, McGill R. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. Journal of the American Statistical Association. 1984;79:531–554.

[16] Forschungsverbund Treat-ION. Register [Internet] [cited 2024 Apr 22]. Available from: https://treat-ion.de/register/.

[17] Section Epilepsy, RWTH Aachen University. EpiReg [Internet] [cited 2024 Jan 2]. Available from: https://github.com/epilepsieukaachen/EpiReg.

[18] Rumbaugh J, Jacobson I, Booch G. Unified Modeling Language Reference Manual, The (2nd Edition). 03212456. 2004.

[19] Ludewig J, Lichter H. Software Engineering : Grundlagen, Menschen, Prozesse, Techniken / Jochen Ludewig, Horst Lichter. 4th edn. Heidelberg: dpunkt.verlag; 2023.