BOAS PRÁTICAS NA GESTÃO DE DADOS PARA INVESTIGAÇÃO CLÍNICA

BEST PRACTICE IN CLINICAL RESEARCH DATA MANAGEMENT

Pier Giorgio Basile

Data Managers and Statisticians Coordinator

Hugo Morgado

eCRF Coordinator

AIBILI association for innovation and biomedical research on light and image

Research Technology Organisation
dedicated to clinical research and development of health technologies

Private non-profit organization, established in 1989
with Public Utility recognition

Interface Centre of the Portuguese Network of the Economy Ministry
unique with focus on Human Health

Clinical research performed according to

- ICH - Good Clinical Practice (GCP)
- General Data Protection Regulation (GDPR)
- National and European regulatory requirements

Certifications

- ISO 9001:2015
- Data Centre: European Clinical Research Infrastructure Network (ECRIN)

Data manufacturing system

Data producers

Data custodians

Data consumers

We define **high-quality data** as data that is fit for the use by data consumers

Data Quality in Contex – Diane M. Strong, Yang W. Lee, Richard Y. Wang

Characteristics of high-quality data

DQ Category	DQ Dimensions
Instrinsic	Accuracy, objectivity, believability, reputation
Accessibility	Accessibility, access security
Contextual	Relevancy, value-added, timeliness, completeness, amount of data
Representational	Interpretability, ease of understanding, coincise representation, consistent representation

Data Quality in Contex – Diane M. Strong, Yang W. Lee, Richard Y. Wang

1. Clinical Protocol

Annotated Protocol

Data manager
eCRF developer
Sponsor
Project Lead Team
Statistician

List of Variables

- type
- unit
- range / decimal
- critical vs non-critical
- mandatory
- properties

Edit Checks

- expression
- target
- action type
- message
- critical vs non-critical
- on-line vs off-line

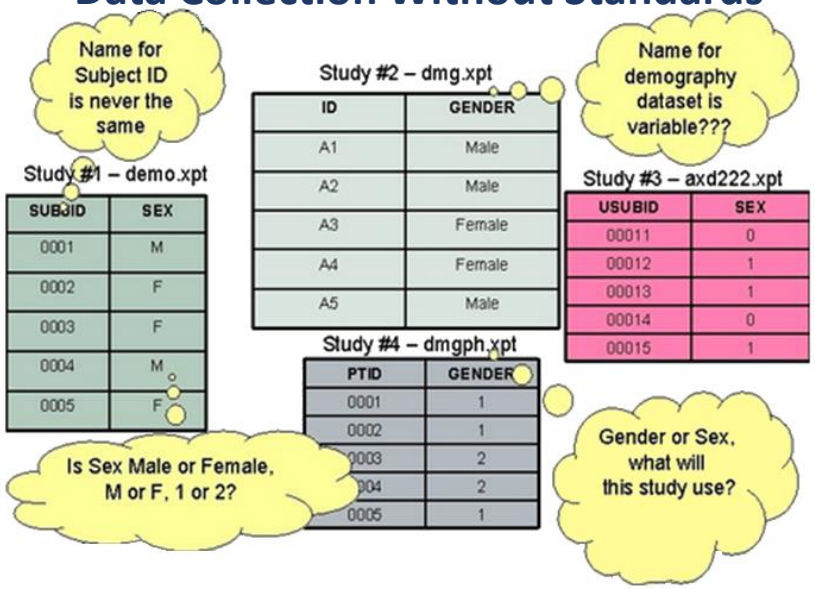
electronic Case Report Form (eCRF)

1. Coding Variables

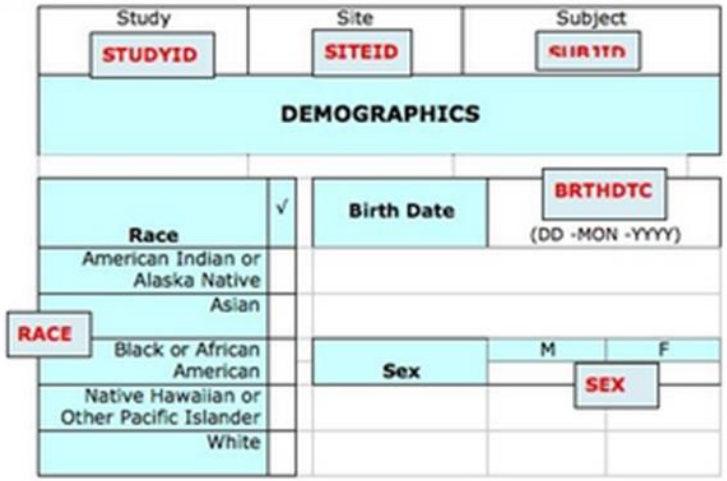
Clinical Data Acquisition Standards Harmonization (CDASH)

- is a foundational CDISC standard
- defines basic standards for the collection of clinical trial data

Data Collection Without Standards



Data Collection With CDASH (paper)



Basic Concepts of CDASH

- Minimal 'core' dataset for clinical research
- Standardize the questions/fields on CRFs
- Standardize the variables and harmonize with SDTM (CDASH is a subset of SDTM)
- Collect data using standard CDISC controlled terminology that maps into SDTM

List of Variables

CRF	Variable Num.	Variable description	Other Properties	Units	Type	Range		No. of Decimals
						Min	Max	
Demography	7	Date of Informed Consent	[DD-MM-YYYY]	NA	Date	NA	NA	NA
Demography	8	Year of birth	[YYYY]	NA	Numeric	1930	1990	NA
Demography	9	Sex	[/Male/Female]	NA	Single-Select	NA	NA	NA
Demography	10	Race	[/Caucasian/ Asian/ African/ other]	NA	Single-Select	NA	NA	NA
Demography	11	If other - specify	If Race = "other"	NA	Text	NA	NA	NA
Vital Signs - Screening	12	Performed	[/Yes/No]	NA	Single-Select	NA	NA	NA
Vital Signs - Screening	13	Reason	IF 'Performed'=No	NA	Text	NA	NA	NA
Vital Signs - Screening	14	Exam date is the same as visit date?	IF 'Performed'=Yes [/Yes / No]	NA	Single-Select	NA	NA	NA
Vital Signs - Screening	15	Exam date	IF 'Exam date is the same as visit date?'=No [DD-MM-YYYY]	NA	Date	NA	NA	NA
Vital Signs - Screening	16	Height	IF 'Performed'=Yes	cm	Numeric	130	210	0
Vital Signs - Screening	17	Weight	IF 'Performed'=Yes	kg	Numeric	30	140	1
Vital Signs - Screening	18	BMI	(CALCULATE AUTOMATIC)	kg/m2	Calculated	12	40	1
Vital Signs - Screening	19	Systolic Blood Pressure	IF 'Performed'=Yes	mmHg	Numeric	70	210	0
Vital Signs - Screening	20	Diastolic Blood Pressure	IF 'Performed'=Yes	mmHg	Numeric	50	120	0
Vital Signs - Screening	21	Heart Rate	IF 'Performed'=Yes	bpm	Numeric	40	300	0

Edit Checks

Rule #	Rule Description	Expression (Always include CRF_NAME.ITEM_NAME)	Target (eCRF item that **triggers** the Rule)	[Action Type]	[Message]	[Expression evaluates to]
1	Visit Date of V2_M6 should be grater than V1_M0 Visit Date	DMDAT (V2_M6) gt DMDAT (V1_M0)	DMDAT	Discrepancy Note Action	Visit Date of V2_M6 should be grater than V1_M0 Visit Date	FALSE
2	Visit Date of V3_M12 should be grater than V2_M6 Visit Date	DMDAT (V3_M12) gt DMDAT (V2_M6)	DMDAT	Discrepancy Note Action	Visit Date of V3_M12 should be grater than V2_M6 Visit Date	FALSE
3	Visit Date of V4_M24 should be grater than V3_M12 Visit Date	DMDAT (V4_M24) gt DMDAT (V3_M12)	DMDAT	Discrepancy Note Action	Visit Date of V4_M24 should be grater than V3_M12 Visit Date	FALSE
4	Date of Informed Consent must be less or equal than V1_M0 Visit Date	DMSTDAT_IC lte DMDAT	DMSTDAT_IC	Discrepancy Note Action	Date of Informed Consent must be less or equal than V1_M0 Visit Date	FALSE
5	Subject enrolled in the Study but none of the 6 month criteria are met.	IEPENROLLEDINOCCUR eq 1 AND IEYN eq 2 AND IEMARETMOCCUR eq 2	IEPENROLLEDINOCCUR	Discrepancy Note Action	The subject is being enrolled in the Study but none of the 6 month criteria are met. Please confirm data.	TRUE
6	Subject enrolled in the Study but at least one of the inclusion criteria is not verified.	IEPENROLLEDINOCCUR eq 1 AND (IEIN001 eq 2 OR IEIN002 eq 2 OR IEIN003 eq 2 OR IEIN004 eq 2 OR IEIN005 eq 2 OR IEIN006 eq 2 OR IEIN007 eq 2 OR IEIN008 eq 2 OR IEIN009 eq 2 OR IEIN010 eq 2)	IEPENROLLEDINOCCUR	Discrepancy Note Action	The subject is being enrolled in the Study but at least one of inclusion criteria is not verified. Please confirm data.	TRUE

2. Data Monitoring

remote

Recruitment period

- Enrolment reports
- Screening failure reports
- Intensive follow-up with sites (missing data, queries, outliers)

Follow-up period

- Drop-outs reports
- Visits performed/missed reports
- Regular follow-up with sites (missing data, queries, outliers)

on-site

Source Data Verification (SDV)

- Trial instructions not followed
- Data entry errors
- Fraud

2. Data Monitoring

remote

Data Managament Plan

- eCRF design, development, and review
- Validation Plan
- Data Quality
- Data Coding
- Clean-file and Database lock
- Allocation of resources
- Ethical aspects
- Data security
- Data access

on-site

Monitoring Manual

- Study procedures
- Study documentation
- Initation Visit
- Monitoring Visits
- Close-out Visits

3. Change Request and Risk Assessment

Collect:

- Reason for change
- Requested change
- Approvals of Sponsor and Statistician (if applicable)

Risk analysis:

- Assess the impact of the change

Can be implemented

Cannot be implemented

Risk Analysis:

Use the checklist below to perform the risk analysis of changes in the CDMA/CRF/GF system

Functional Compliance	Y/N	Technical Compliance	Y/N
Impact on Internal or External Regulation		Impact on integrity tests/validation scripts	
User will probably resist to change		Impact on storage size	
Requires protocol amendment		Impact on Import and Export scripts	
Re-training of users		Backup and production information consistency	
Documentation review		Technical Revalidation (IQ/PQ) needed	
Communication review		Operational Revalidation (OQ) needed	
Ethical Questions		Impact on Third Party integration software	
Impact on data already entered		Impact on Database or interfaces structure"	
<..others..>		<..others..>	

4. Data Coding

Coding MH, AE, e CM

Medical History (MH) -> MedDRA

Adverse Events (AE) -> MedDRA

In the late 1990s, the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) developed MedDRA, a rich and highly specific standardised medical terminology to facilitate sharing of regulatory information internationally for medical products used by humans. (https://www.meddra.org/)

Concomitant Medication & Non-drug therapies -> ATC/DDD Index

The Anatomical Therapeutic Chemical (ATC) classification system and the Defined Daily Dose (DDD) as a measuring unit have become the gold standard for international drug utilization monitoring and research. The ATC/DDD system is a tool for exchanging and comparing data on drug use at international, national or local levels. (https://www.whocc.no/)

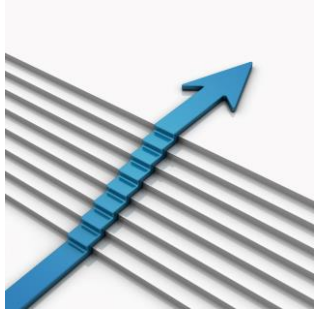
4. Data Coding

Coding MH, AE, e CM

Term in free-text - eCRF	Low Level Term (LLT) MedDRA	*Issue*
Cataract surgery	Cataract operation 10063797	*use of the right term*
Iridodialise	Iridodialysis 10022942	*language*
Geographic athrophy	Geographic atrophy 10063947	*misprint*

→ Standardization
Harmonization

→ Accessibility
Interpretability
Ease of understanding
Value-added



5. Database cleaning & Lock

Data manager
eCRF developer
Sponsor
Project Lead Team
Statistician

Organizations have different approaches

- **Accept no risk:** clean everything to the maximum extent possible.
- **Accept more risk:** don't clean. Do an assessment audit and fix findings that impact the analysis.
- **Mitigate risk:**
  - Clean the major independent, dependent variables and covariates.
  - Clean to reduce outliers.
  - Review draft tables and listings for obvious

Check-list for Database Lock

- ✓ *all queries resolved*
- ✓ *all expected subjects entered*
- ✓ *check for double subjects is negative*
- ✓ *check for blank data is negative*
- ✓ *coding completed*
- ✓ *SAE reconciliation completed*
- ✓ *final quality check completed*
- ✓ *database access rights removed*

6. Database transfer and storage

To be sent:

Final data
- Final database

Metadata
- Queries database
- Codebook
- Guidelines for reading the database and the codebook

Report
- Database Clean Report

How:

- Not-processed data, if possible
- Format previously agreed with Sponsor and Statistician
- Encryption (example: 7Zip in AES256 with password)
- Security (example: OneDrive with password and expiration date)

Retention:
- Long-Term Data Storage

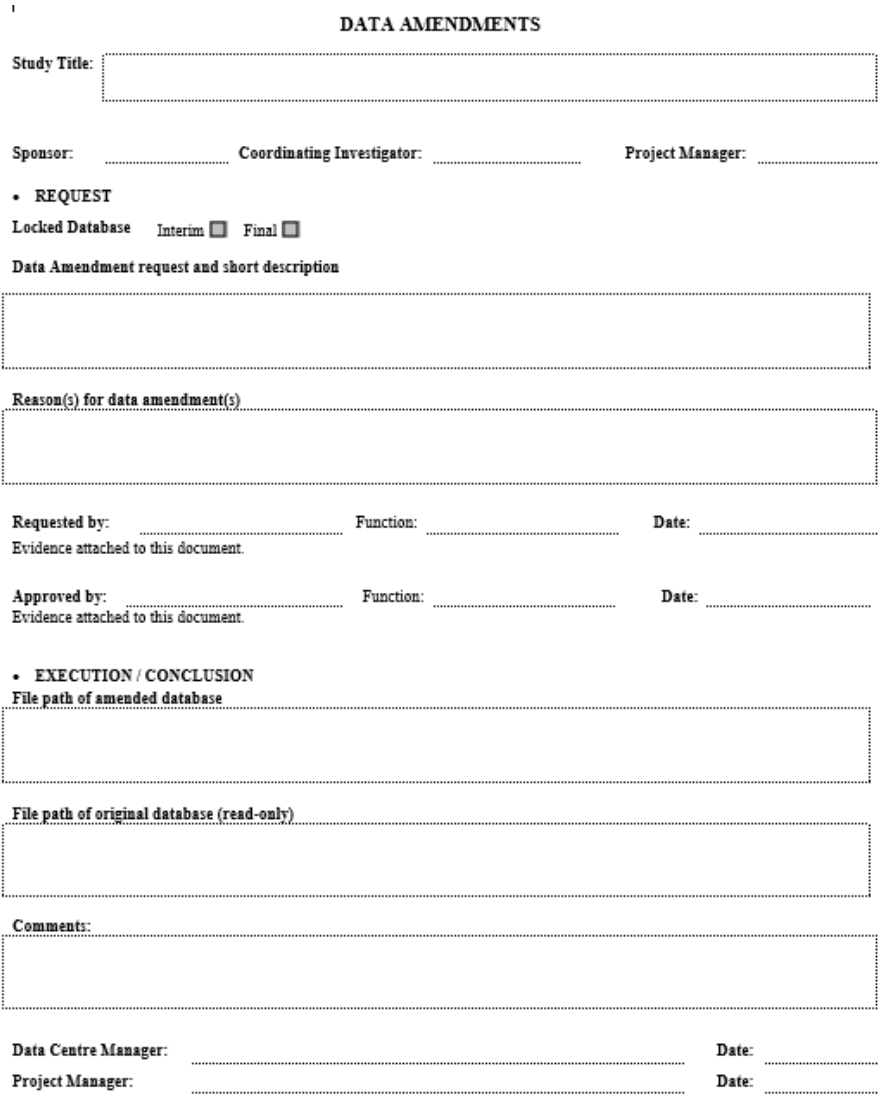
7. Post-lock data amendment

Collect:

- Data amendment request, and short descirption.
- Reason for data amendment.
- File path of amendend database.
- File path of original database.
- Approvals of Sponsor and Statistician.

Ensure that:

- The data have been amended according to the request. ✓



DATA AMENDMENTS

Study Title:

Sponsor: Coordinating Investigator: Project Manager:

- REQUEST

Locked Database Interim ☐ Final ☐

Data Amendment request and short description

Reason(s) for data amendment(s)

Requested by: Function: Date:
Evidence attached to this document.

Approved by: Function: Date:
Evidence attached to this document.

- EXECUTION / CONCLUSION

File path of amended database

File path of original database (read-only)

Comments:

Data Centre Manager: Date:

Project Manager: Date:

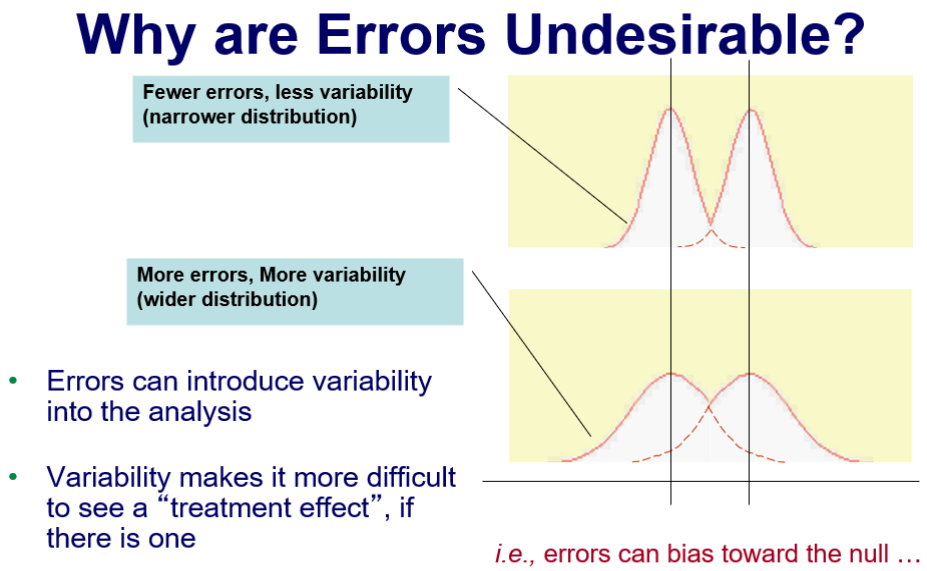
Typical Impacts

Inaccurate data
- errors can introduce variability into the analysis.
- variability makes it more difficult to see a “treatment effect”, if there is one.

Operational impacts
- Lowered customer satisfaction
- Increased cost
- Lowered employee satisfaction

Typical impacts
- Poorer decision making
- More difficult to implement data warehouses
- More difficult to reengineer
- Increased organizational mistrust

Strategic impacts
- More difficult to set and execute strategy
- Contribute to issues of data ownership
- Compromise ability to align organizations
- Divert management attention



Data Quality in Clinical Research
Society for Clinical Data Management

50% to 80% of computerized criminal records in the U.S. were found to be inaccurate, incomplete, or ambiguous. The social and economic impact of poor-quality data costs billions of dollars

Data Quality in Context – Diane M. Strong, Yang W. Lee, and Richard Y. Wang – Communications of the ACM, May 1997 Vol 40, nº5

Thank you for your attention

CONTACTS:

PIER GIORGIO BASILE
Edifício Prof. Doutor José Cunha-Vaz
Azinhaga Sta. Comba, Celas
3000-548 Coimbra
Portugal

Phone: +351 239 480 142
E-mail: pgbasile@aibili.pt

HUGO NOGUEIRA MORGADO
Edifício Prof. Doutor José Cunha-Vaz
Azinhaga Sta. Comba, Celas
3000-548 Coimbra
Portugal

Phone: +351 239 480 150
E-mail: hmorgado@aibili.pt

FITEC Fundo de Inovação, Transferência de Tecnologia e Economia Circular
PROGRAMA INTERFACE
Programa Nacional de Reformas
REPÚBLICA PORTUGUESA