

Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern

M. Nonnemacher¹, D. Weiland¹, J. Stausberg²

¹ Institut für Medizinische Informatik, Biometrie und Epidemiologie (IMIBE)
Medizinische Fakultät der Universität Duisburg-Essen
Universitätsklinikum Essen

² Institut für Medizinische Informationsverarbeitung, Biometrie und
Epidemiologie (IBE)
Ludwig-Maximilians-Universität München

Ein Projekt für die vernetzte
medizinische Forschung im
Namen der TMF

TMF – Technologie- und Methodenplattform
für die vernetzte medizinische Forschung e.V.



Gefördert vom



Bundesministerium
für Bildung
und Forschung

Inhalt

- Einführung
- Projektziele
- Vorgehensweise
- Indikatoren für Datenqualität
- Datenqualitätsscore
- Empfehlungen für Source Data Verification und Feedback
- Workflow

Probleme von Kohortenstudien und Registern

- Datenmenge
 - Hohe Anzahl eingeschlossener Personen
 - Hohe Anzahl erhobener Variablen
- Datenübernahme
 - Verschiedene Datenquellen (z.B. Routinedaten)
 - Vorzugsweise über Anwendungsschnittstellen
- Bestimmung, Sicherung und Verbesserung der Datenqualität notwendig
- Vollständiger Originaldatenabgleich (Source Data Verification = SDV) nicht praktikabel

Projektziele

- Effiziente Sicherung hoher Datenqualität
 - Bestimmbare (hohe) Datenqualität
 - Kalkulierbare (bezahlbare) Kosten
- Anpassung der Datenmanagement-Aktivitäten an die Datenqualität
 - Adaption an die Qualität der gespeicherten Daten
 - Adaption an die Ergebnisse einer vorherigen SDV

Projektergebnisse

- Leitlinie für das adaptive Management von Datenqualität in Kohortenstudien und Registern
- Software zur Unterstützung der Leitlinienanwendung

Framework of Procedures for the Assurance of Data Quality in Medical Registries

Central coordinating centre	Local sites
Prevention during set up and organisation of registry	
At the onset of the registry compose minimum set of necessary data items define data & data characteristics in data dictionary draft a data collection protocol define pitfalls in data collection compose data checks create user friendly case record forms create quality assurance plan In case of new participating sites perform site visit train new participants Continuously motivate participants communicate with local sites In case of changes (e.g., in data set) adjust forms, software, data dictionary, protocol, training material, etc. communicate with local sites	At the onset of participating in the registry assign a contact person check developed software for data entry and for extraction check reliability and completeness of extraction sources standardise correction of data items Continuously train (new) data collectors motivate data collectors make data definitions available place date & initials on completed forms keep completed case record forms data collection close to the source and as soon as possible use the registry data for local purposes In case of changes (e.g., in data set) adjust data dictionary, forms, software, etc. communicate with data collectors
Detection during data collection	
During import of data into the central database perform automatic data checks Periodically and in case of new participants perform site visits for data quality audit (registry data ↔ source data) and review local data collection procedures Periodically check inter- and intraobserver variability perform analyses on the data	Continuously visually inspect completed forms perform automatic data checks check completeness of registration
Actions for quality improvement	
After data import and data checks provide local sites with data quality reports control local correction of data errors After data audit or variability test give feedback of results and recommendations resolve causes of data errors	After receiving quality reports check detected errors correct inaccurate data & fill in incomplete data resolve causes of data errors After receiving feedback implement recommended changes communicate with personnel

Vorgehensweise

- Workshop „Medizinische Daten in Forschung und Versorgung“
- Analyse existierenden Wissens
 - Literaturreview: 146 Dokumente
 - Expertenbefragungen
- Entwicklung von Verfahren für
 - Messung und Beurteilung von Datenqualität
 - Adaption von Häufigkeit, Umfang und Tiefe der SDV an die Datenqualität
- Partner: 4 Kompetenznetze in der Medizin
- Erstellung der konsentierten Leitlinie als Dokument
- Implementierung der Leitlinie als Web-Service

Beispiel für Datenqualitätsindikator 1

Beschreibung	Anzahl fehlender Werte bei mandatorischen Datenelementen
Begriffsdefinitionen	---
Typ	Ergebnisqualität
Quelle	Aronsky und Haug 2000, Gaus 2003, ... [weitere Quellen]
Bezug	Die Berechnung des Indikators ist nur für <u>mandatorische</u> Datenelemente sinnvoll.
Alternative Definition	---
Anmerkungen	<p>Der Indikator ist wichtig für die korrekte Schätzung von Inzidenzen und Prävalenzen. Der Anteil der Fälle mit vollständigen Daten ist für spätere multivariate Analysen von großer Bedeutung.</p> <p>Ein entsprechender Indikator ist auch für optionale Datenelemente vorhanden. Die separate Behandlung von mandatorischen und optionalen Datenelementen wird von Aronsky und Haug vorgeschlagen, da jeweils unterschiedliche Fehlerraten tolerabel sind (Aronsky und Haug 2000).</p>

Beispiel für Datenqualitätsindikator 2

Zähler	Anzahl fehlender Werte bei mandatorischen Datenelementen
Nenner	Anzahl überprüfter Werte bei mandatorischen Datenelementen
Subkategorien	---
Berechnung	1) Suche nach fehlenden Werten bei mandatorischen Datenelementen 2) Berechnung von Zähler, Nenner und Rate Alternative Berechnungsmöglichkeit: Anzahl fehlender Werte pro Patient/Person
Interpretation	Je höher die Rate, desto schlechter ist die Datenqualität. Welche Rate noch akzeptabel ist, hängt von der vorgesehenen Datennutzung ab.
Einflussfaktoren	Einführung von separaten Kategorien für „not applicable“, „not done“

Indikatoren für Datenqualität 1

Ebene Plausibilität (Strukturqualität)

Qualitätsindikator	Schwellenwert für Auffälligkeit	Gewicht
Übereinstimmung mit Vorwert	> 5 %	1
Konkordanz	> 5 %	1
Widerspruchsfreiheit	> 5 %	3
Werte Verteilung innerhalb und zwischen Zentren	> 10 %	1
Untersuchungen am Wochenende oder an gesetzlichen Feiertagen	> 2 %	1
Anzahl fehlender Werte bei optionalen Datenelementen	> 5 %	1
Ausreißer bei stetigen Datenelementen	> 10 %	2
Erlaubte Werte bei qualitativen Datenelementen	> 0 %	1
Bevorzugung bestimmter Endziffern	> 5 %	1
Nachweis bekannter Korrelationen	< 10 %	1

Indikatoren für Datenqualität 2

Ebene Organisation (Prozessqualität)

Qualitätsindikator	Schwellenwert für Auffälligkeit	Gewicht
Aktualität der gespeicherten Daten	> 10 %	6
Anzahl Dubletten im Datenbestand	> 5 %	4
Rekrutierungsrate	< 90 %	5
Anzahl Patienten, die vorzeitig aus dem Register ausscheiden (Drop-out-Rate)	> 1 %	5
Fachliche Qualifikation des für die Datenerhebung und – erfassung im Register zuständigen Personals in den Zentren	< 100 %	6
Anzahl Synonyme	> 5 %	4
Anzahl Homonymé	> 2 %	4

Indikatoren für Datenqualität 3

Ebene Richtigkeit (Ergebnisqualität)

Qualitätsindikator	Schwellenwert für Auffälligkeit	Gewicht
Genauigkeit der Registerdaten	< 95 %	8
Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Datenelemente	> 5 %	9
Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Beobachtungseinheiten	> 5 %	8
Vollständigkeit der Registerdaten	< 95 %	7
Übereinstimmung mit dem Studienplan	> 5 %	7
Anzahl fehlender Werte bei mandatorischen Datenelementen	> 5 %	7
Repräsentativität der Registerdaten	< 100 %	7

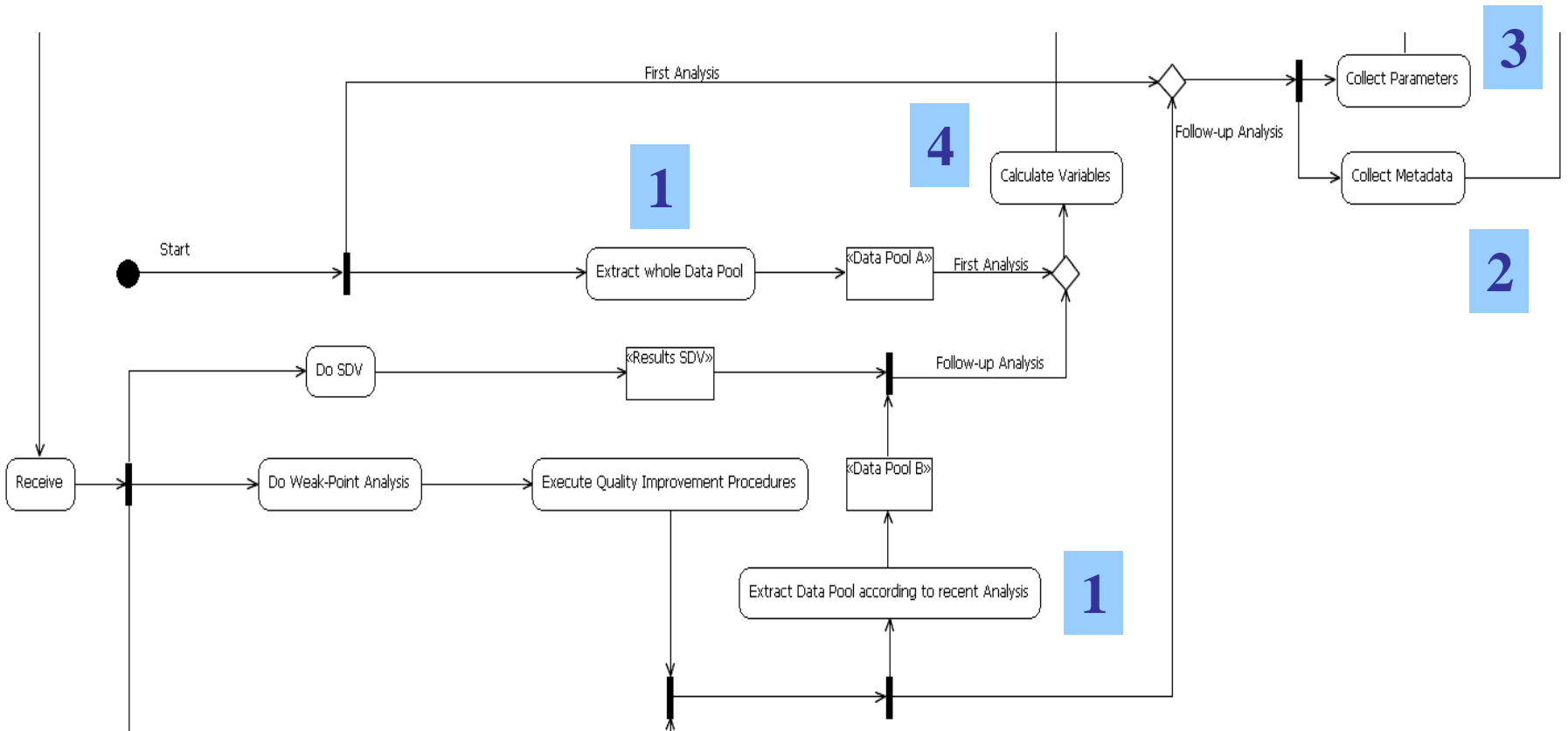
Berechnung des Qualitätsscores

- Für jeden Qualitätsindikator
 1. Schwellenwert für auffällig/unauffällig
 2. Relatives Gewicht
- Berechnung des Qualitätsscores
 1. Wert 1 für jeden unauffälligen Indikator, Wert 0 für jeden auffälligen Indikator
 2. Multiplikation des Wertes mit dem relativen Gewicht
 3. Addition der Werte für alle bestimmmbaren Indikatoren
 4. Transformation auf eine Skala von 0 bis 100
 5. Bestimmung der Datenqualitätskategorie
 - 0-20 sehr schlecht
 - 21-40 schlecht
 - 41-60 moderat
 - 61-80 gut
 - 81-100 sehr gut

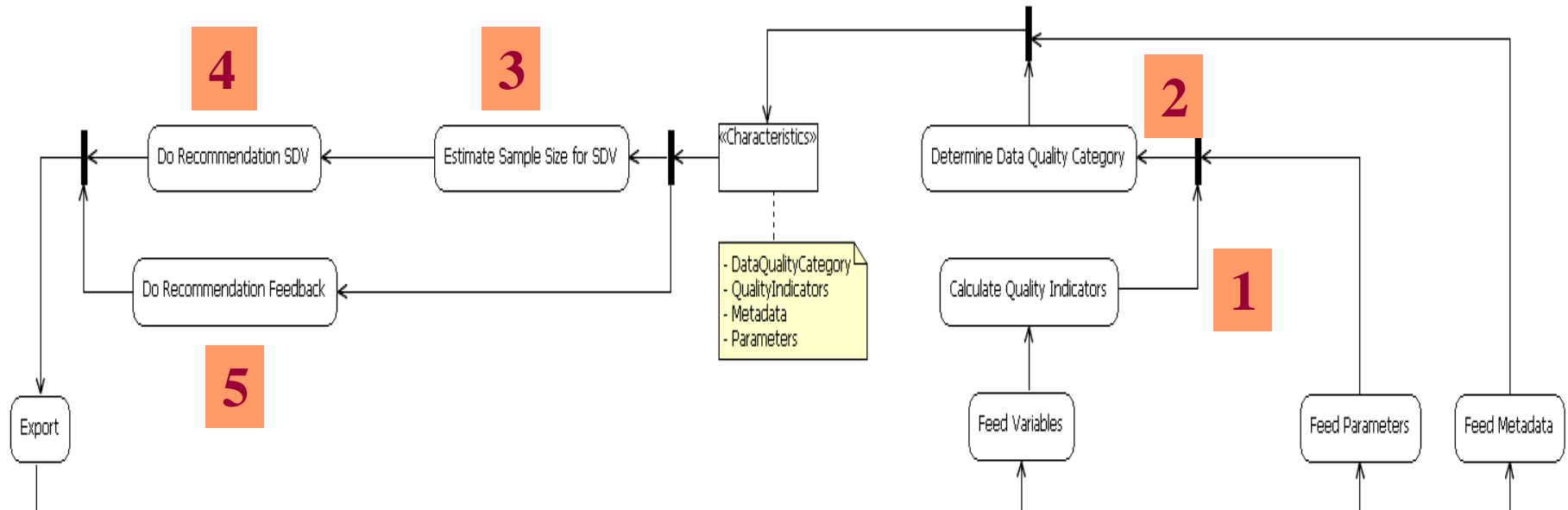
Empfehlungen

- **Source Data Verification (SDV)**
 - Fallzahlschätzung für Anzahl der Patienten
 - Fallzahlschätzung für Anzahl der Datenelemente
 - Häufigkeit der Besuche in den Studienzentren
- **Feedback über Datenqualität (je Zentrum)**
 - Eigene Ergebnisse, oder
 - Eigene Ergebnisse + anonymisierte Vergleichsdaten der anderen Zentren, oder
 - Eigene Ergebnisse + offene Vergleichsdaten der anderen Zentren
- **Zeitpunkt der nächsten Leitlinienanwendung**

Workflow: Vorbereitung



Workflow: Anwendung der Leitlinie



Software: Ergebnisse eines fiktiven Erhebungszentrums (Ausschnitt)

Leitlinie zum adaptiven Management von Datenqualität - Microsoft Internet Explorer

Adresse: https://imibe03ux.uk-essen.de/guideline/calc_feedback_site.faces

TMF Telematikplattform für Medizinische Forschungsnetze e. V.

Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern

Vorhaben: Test Register
Analyse: 1
Leitlinie: Version 1.0

Erhebungszentrum: Universitätsklinik Essen

Allgemeine Angaben

Beschreibung	Register des Kompetenznetzes XYZ
Angestrebtes Niveau der Datenqualität	gut
Vergleich der Rate mit dem Schwellenwert durch Konfidenzintervall	ja
Analyseperiode	0 Jahre 3 Monate 0 Tage
Gesamtscorewert*	59.0
Beurteilung	moderat

Ebene Plausibilität

Qualitätsindikator	Rate	Schwellenwert	Konfidenzintervall	Auffällig*	Gewichtung
Übereinstimmung des Ergebnisses einer Messung oder Untersuchung mit dem Vorwert	0.3%	> 5%	0.1% - 0.9%	Nein	1
Konkordanz	2.7%	> 5%	1.8% - 3.9%	Nein	1
Widerspruchsfreiheit	29.6%	> 5%	26.8% - 32.5%	Ja	3
Werteverteilungen innerhalb und zwischen Zentren	2.6%	> 10%	0.5% - 7.3%	Nein	1

Abmelden
Bitte vergessen Sie nicht, sich auszuloggen
Logout

Software: Empfehlungen zur SDV für ein fiktives Erhebungszentrum

Leitlinie zum adaptiven Management von Datenqualität - Microsoft Internet Explorer

Adresse https://imibe03ux.uk-essen.de/guideline/calc_sdv_site.faces

TMF Telematikplattform für Medizinische Forschungsnetze e. V.

Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern

[TMF](#) [Startseite](#) [Kontakt](#) [Info](#) [Impressum](#)

Vorhaben: Test Register
Analyse: 1
Leitlinie: Version 1.0

Erhebungszentrum: Universitätsklinik Essen

Abmelden
Bitte vergessen Sie nicht, sich auszuloggen
[Logout](#)

Allgemeine Angaben	
Beschreibung des Vorhabens	Register des Kompetenznetzes XYZ
Analyseperiode	0 Jahre 3 Monate 0 Tage
Umfang der SDV	
Fehlerrate Beobachtungseinheiten	0,325
Nichtadjustierte Fallzahl	937
Gesamtzahl der Beobachtungseinheiten	750
Adjustierte Fallzahl	417
Tiefe der SDV	
Fehlerrate Datenelemente	0,043
Nichtadjustierte Fallzahl	176
Gesamtzahl der Datenelemente	250
Adjustierte Fallzahl	104
Frequenz der SDV	
Kapazität für SDV pro Besuch	30
Anzahl der Besuche	14

Kontinuierliche Verbesserung der Datenqualität

Überprüfung

```
OP-Dokumentation
Abteilung: UG2 Benutzer: DR. STAUBERG, DMBS Klinikum DHZ Essen 1
Name: Hubertmann, Karl 966 Datum: 08/03/10-40
Station: H2 Zimmernr.: Bett: FallNr.: 1458 Aufn.Datum: 08/07/09

OP-Termin : 050907 OP-Zahl: A1 Lagerung: NUCC
gepl. Eingr. : NOTFALL-EINGRIFF
Operateur : RADI Assistent: KOL
ausw.Mit. : Beiratung: (J/N)
Op-Beginn : 1000 Op-Schnitt: 1015 Op-Nacht: 1045 Op-Ende : 1100
Eingriff : Cholezystektomie einfach laparoskopisch ICD: 581111
Cholezystektomie einfach laparoskopisch
Ohne laparoskopische Revision der Gallengänge
Operationart: L. Reizg./Bündel (R/A) = Fallpausch.: 1204 Sondgltr.: 1212
postop.Diagn.: Cholezystolithiasis mit sonstiger Cholez ICD: 5741
Cholezystolithiasis mit sonstiger Cholez Zus:
Diagnoseart : A klinisch
Bes.heiten :

Komplikat. :
*** Bitte überprüfen ***
Dia/OP-KZ 2 Team Weitere Eingr. Zusl. Verbr. Nachs. Speichern Zurück
KOPEDOAC F1 F2 DATFREI F6 F8 F9 F12
```

Angepasste Empfehlungen

Verbesserung der Datenqualität

Durchführung: SDV, Analyse

Schriftenreihe der Telematikplattform
für Medizinische Forschungsnetze

M. Nonnemacher
D. Weiland | J. Stausberg



Datenqualität in der medizinischen Forschung

Leitlinie zum adaptiven Management
von Datenqualität in
Kohortenstudien und Registern

 Medizinisch Wissenschaftliche Verlagsgesellschaft

Leitlinie + Software
verfügbar über

<http://www.tmf-ev.de>

Buch auch verfügbar über
Buchhandel

ISBN: 978-3-939069-36-2

Danksagung

Die Entwicklung der Leitlinie ist ein Projekt für die vernetzte medizinische Forschung im Namen der TMF und wurde vom Bundesministerium für Bildung und Forschung gefördert. Wir danken unseren Projektpartnern von den Kompetenznetzen HIV/AIDS, Demenzen, Angeborene Herzfehler und Parkinson für ihre Kooperation.

Vielen Dank für Ihre Aufmerksamkeit!