

Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern

M. Nonnemacher, D. Weiland, M. Neuhäuser,
J. Stausberg

Institut für Medizinische Informatik, Biometrie und Epidemiologie,
Medizinische Fakultät, Universität Duisburg-Essen

Ein Projekt für die vernetzte
medizinische Forschung im
Namen der TMF

Telematikplattform für
Medizinische Forschungsnetze e. V.



Gefördert vom

Bundesministerium
für Bildung
und Forschung



TMF e.V.

- Telematikplattform für medizinische Forschungsnetze e.V.
 - Meta-Organisation medizinischer Forschungsverbände:
 - Kompetenznetze in der Medizin
 - Koordinierungszentren für Klinische Studien
 - Weitere Verbände
 - Ziel: Verbesserung der Organisation und Infrastruktur für die vernetzte medizinische Forschung
 - Förderung durch BMBF

Probleme von Kohortenstudien und Registern



- Datenmenge
 - Hohe Anzahl eingeschlossener Patienten
 - Hohe Anzahl von erhobenen Variablen
- Datenübernahme
 - Verschiedene Datenquellen (z.B. Routinedaten)
 - Vorzugsweise über Anwendungsschnittstellen
- Bestimmung und Sicherung der Datenqualität notwendig
- Vollständiger Originaldatenabgleich (Source Data Verification = SDV) zu teuer



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
- Guideline für das adaptive Management von Datenqualität
- 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 <ul style="list-style-type: none">compose minimum set of necessary data itemsdefine data & data characteristics in data dictionarydraft a data collection protocoldefine pitfalls in data collectioncompose data checkscreate user friendly case record formscreate quality assurance plan	At the onset of participating in the registry <ul style="list-style-type: none">assign a contact personcheck developed software for data entry and for extractioncheck reliability and completeness of extraction sourcesstandardise correction of data items
In case of new participating sites <ul style="list-style-type: none">perform site visittrain new participants	Continuously <ul style="list-style-type: none">train (new) data collectorsmotivate data collectorsmake data definitions availableplace date & initials on completed formskeep completed case record formsdata collection close to the source and as soon as possibleuse the registry data for local purposes
Continuously <ul style="list-style-type: none">motivate participantscommunicate with local sites	In case of changes (e.g., in data set) <ul style="list-style-type: none">adjust data dictionary, forms, software, etc.communicate with data collectors
In case of changes (e.g., in data set) <ul style="list-style-type: none">adjust forms, software, data dictionary, protocol, training material, etc.communicate with local sites	
Detection during data collection	
During import of data into the central database <ul style="list-style-type: none">perform automatic data checks	Continuously <ul style="list-style-type: none">visually inspect completed formsperform automatic data checkscheck completeness of registration
Periodically and in case of new participants <ul style="list-style-type: none">perform site visits for data quality audit (registry data ↔ source data) and review local data collection procedures	
Periodically <ul style="list-style-type: none">check inter- and intraobserver variabilityperform analyses on the data	
Actions for quality improvement	
After data import and data checks <ul style="list-style-type: none">provide local sites with data quality reportscontrol local correction of data errors	After receiving quality reports <ul style="list-style-type: none">check detected errorscorrect inaccurate data & fill in incomplete dataresolve causes of data errors
After data audit or variability test <ul style="list-style-type: none">give feedback of results and recommendationsresolve causes of data errors	After receiving feedback <ul style="list-style-type: none">implement recommended changescommunicate 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 Guideline als Dokument
- Implementierung der Guideline als Web-Service

Indikatoren für Datenqualität 1



Ebene Plausibilität

- Übereinstimmung mit Vorwert
- Konkordanz
- Widerspruchsfreiheit
- Werteverteilung innerhalb und zwischen Zentren
- Nachweis bekannter Korrelationen
- Untersuchungen am Wochenende oder an gesetzlichen Feiertagen
- Anzahl fehlender Werte bei optionalen Datenelementen
- Ausreißer bei stetigen Datenelementen
- Erlaubte Werte bei qualitativen Datenelementen
- Bevorzugung bestimmter Endziffern

Ebene Organisation

- Aktualität der gespeicherten Daten
- Anzahl Dubletten im Datenbestand
- Rekrutierungsrate
- Anzahl Patienten, die vorzeitig aus dem Register ausscheiden (Drop-out-Rate)
- Fachliche Qualifikation des für die Datenerhebung und -erfassung im Register zuständigen Personals in den Zentren
- Anzahl Synonyme
- Anzahl Homonymé



Indikatoren für Datenqualität 2

Ebene Richtigkeit

- Genauigkeit der Registerdaten
- Übereinstimmung der Registerdaten mit den Originaldaten
- Vollständigkeit der Registerdaten
- Übereinstimmung mit dem Studienplan
- Anzahl fehlender Werte bei mandatorischen Datenelementen
- Repräsentativität der Registerdaten

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

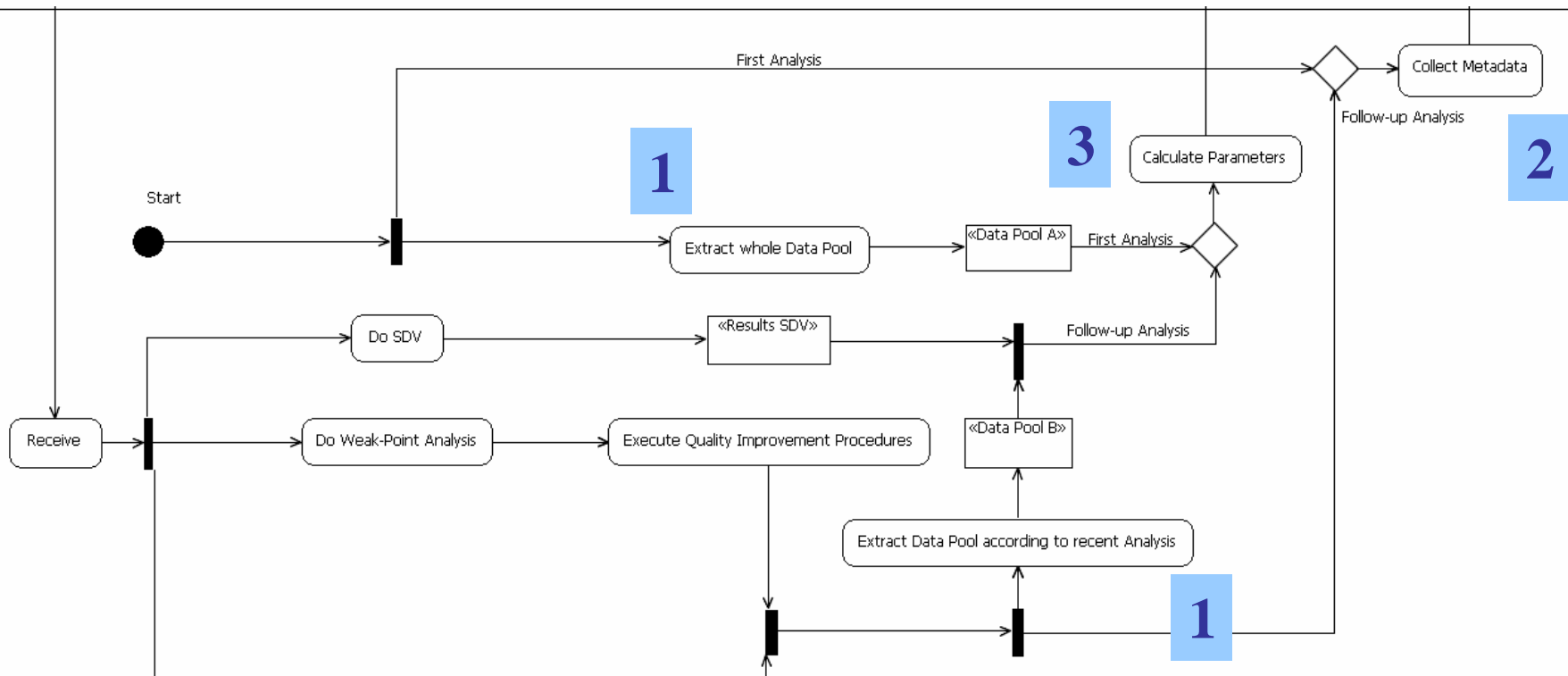


Empfehlungen

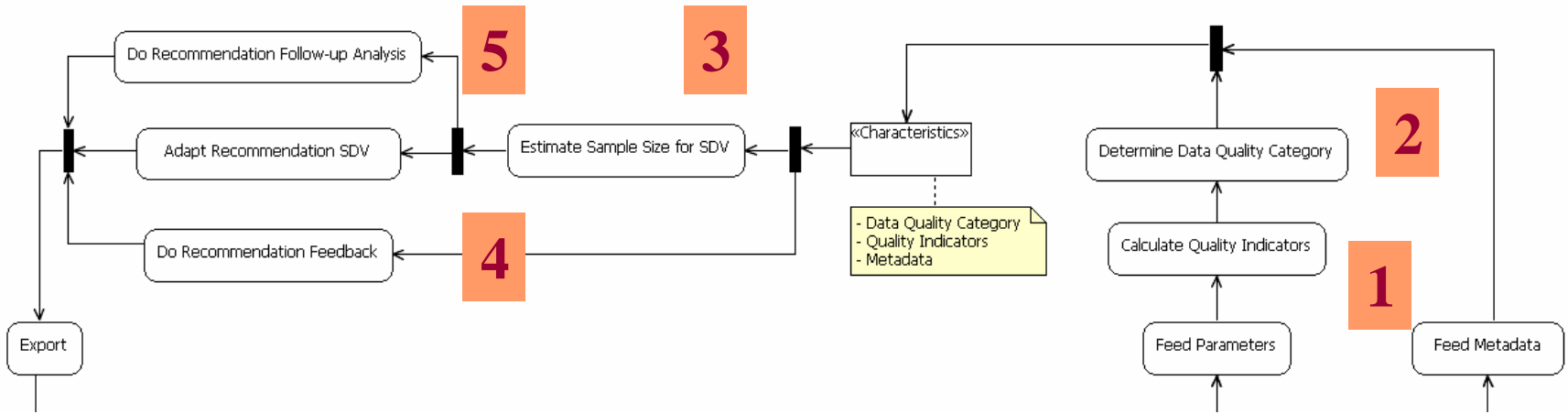
- **Source Data Verification**
 - Fallzahlschätzung für Anzahl der Patienten
 - Fallzahlschätzung für Anzahl der Datenelemente
 - Häufigkeit der Besuche in den Studienzentren
- **Feedback über Datenqualität (je Zentrum)**
 - Eigene Ergebnisse
 - Eigene Ergebnisse + anonymisierte Vergleichsdaten der anderen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



Workflow: Vorbereitung



Workflow: Anwendung der Leitlinie



Workflow: Ausführung der Empfehlungen





Implementierung

- Internetbasierter Dienst
 - Mandantenfähigkeit
 - Sichere Verbindung via https
- Applikations-Framework
 - WWW-Server (Apache)
 - Application server (Tomcat)
 - Servlets
 - Java Server Faces





Leitlinie verfügbar über
info@tmf-ev.de

Vielen Dank für Ihre Aufmerksamkeit!