

Wie lässt sich Datenqualität messen?

Eine Leitlinie der Telematikplattform für Medizinische Forschungsnetze

M. Nonnemacher, D. Weiland, J. Stausberg

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

Inhalt

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

Einführung

- Stark wachsender Umfang elektronisch gespeicherter Gesundheitsdaten
 - Steigende Erwartungen an die Nutzbarkeit dieser Daten (auch über ihren primären Verwendungszweck hinaus)
- Fragen
- Feststellung der Datenqualität (Wie kann man sie messen?)
 - Steuerung / Verbesserung der Datenqualität

Projektziele 1

- 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 eines vorherigen Originaldatenabgleichs (Source Data Verification = SDV)

Projektziele 2

- Leitlinie 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 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

Qualitätsindikator	Schwellenwert für Auffälligkeit	Gewicht
Übereinstimmung mit Vorwert	> 5 %	1
Konkordanz	> 5 %	1
Widerspruchsfreiheit	> 5 %	3
Werteverteilung 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

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

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

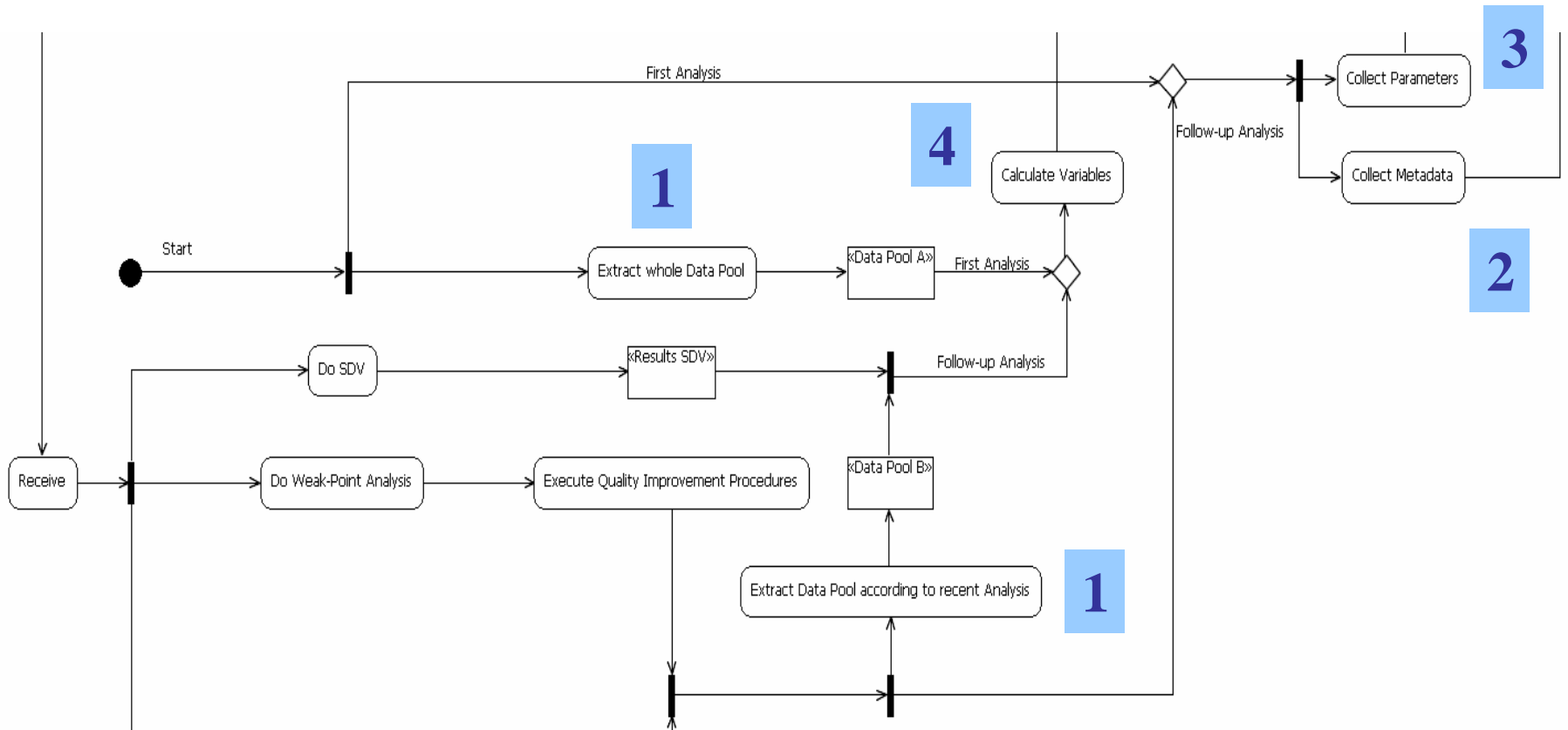
Berechnung des Qualitätsscores

- 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

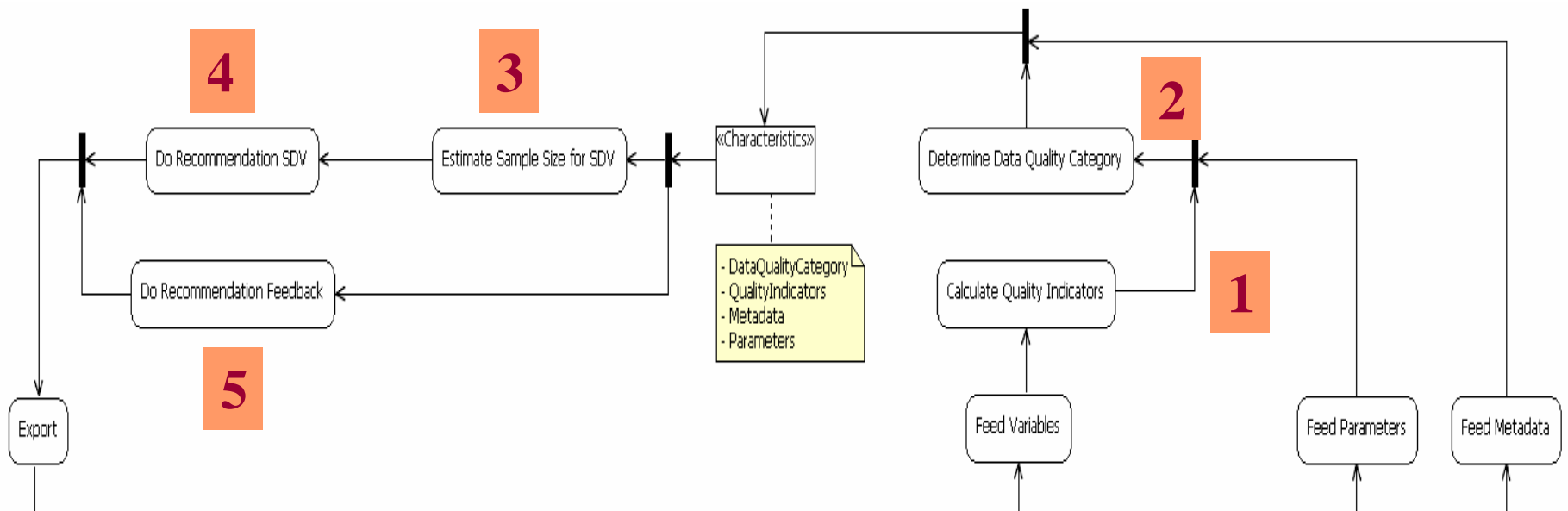
Empfehlungen

- **Source Data Verification**
 - 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**

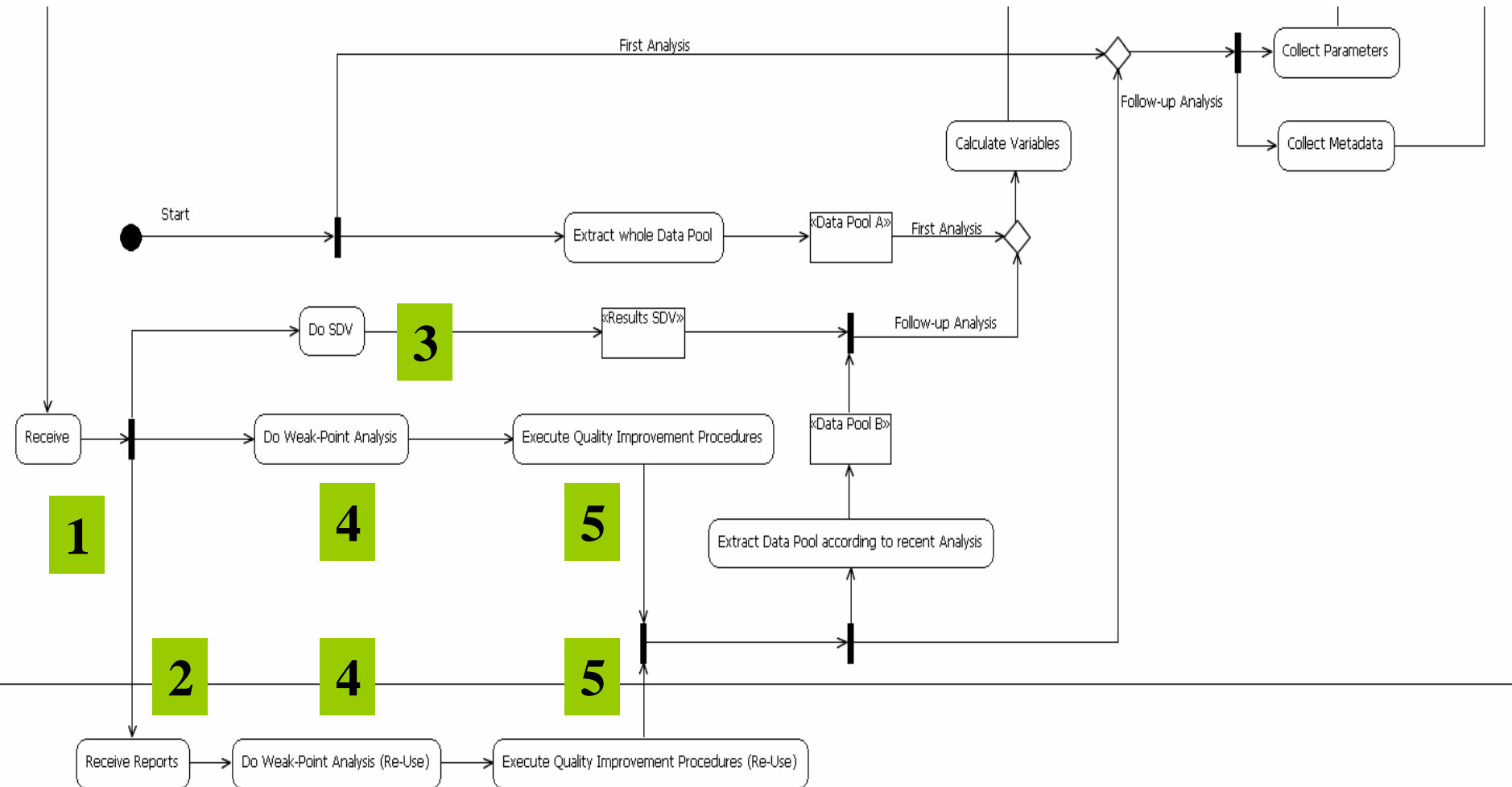
Workflow: Vorbereitung



Workflow: Anwendung der Leitlinie



Workflow: Ausführung der Empfehlungen



Software: Screenshot

Leitlinie zum adaptiven Management von Datenqualität - Microsoft Internet Explorer

Adresse Wechseln zu Links

TMF Telematikplattform für Medizinische Forschungsnetze e. V.

Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern

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

Vorhaben
Analysen
Upload
Metadaten
Parameter
Variablen
Ergebnis

Vorhaben: Test Register
Analyse: 0
Leitlinie: Version 1.0

Analyse auswählen
Analyse:

Neue Analyse erstellen
basierend auf:

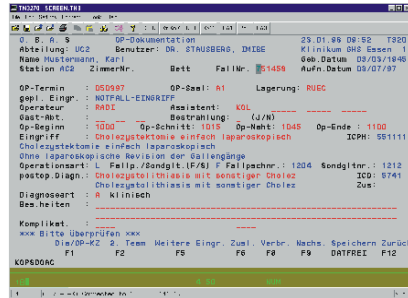
Analyse löschen
Analyse löschen:

Abmelden
Bitte vergessen Sie nicht, sich auszuloggen

Fertig

Kontinuierliche Verbesserung der Datenqualität

Überprüfung



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 verfügbar über
info@tmf-ev.de

oder im 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!