

Datenqualität in epidemiologischen Kohortenstudien

Registertage TMF, 07.05.2019

Birgit Schauer, Carsten Oliver Schmidt

Universitätsmedizin Greifswald

ICM-SHIP-KEF

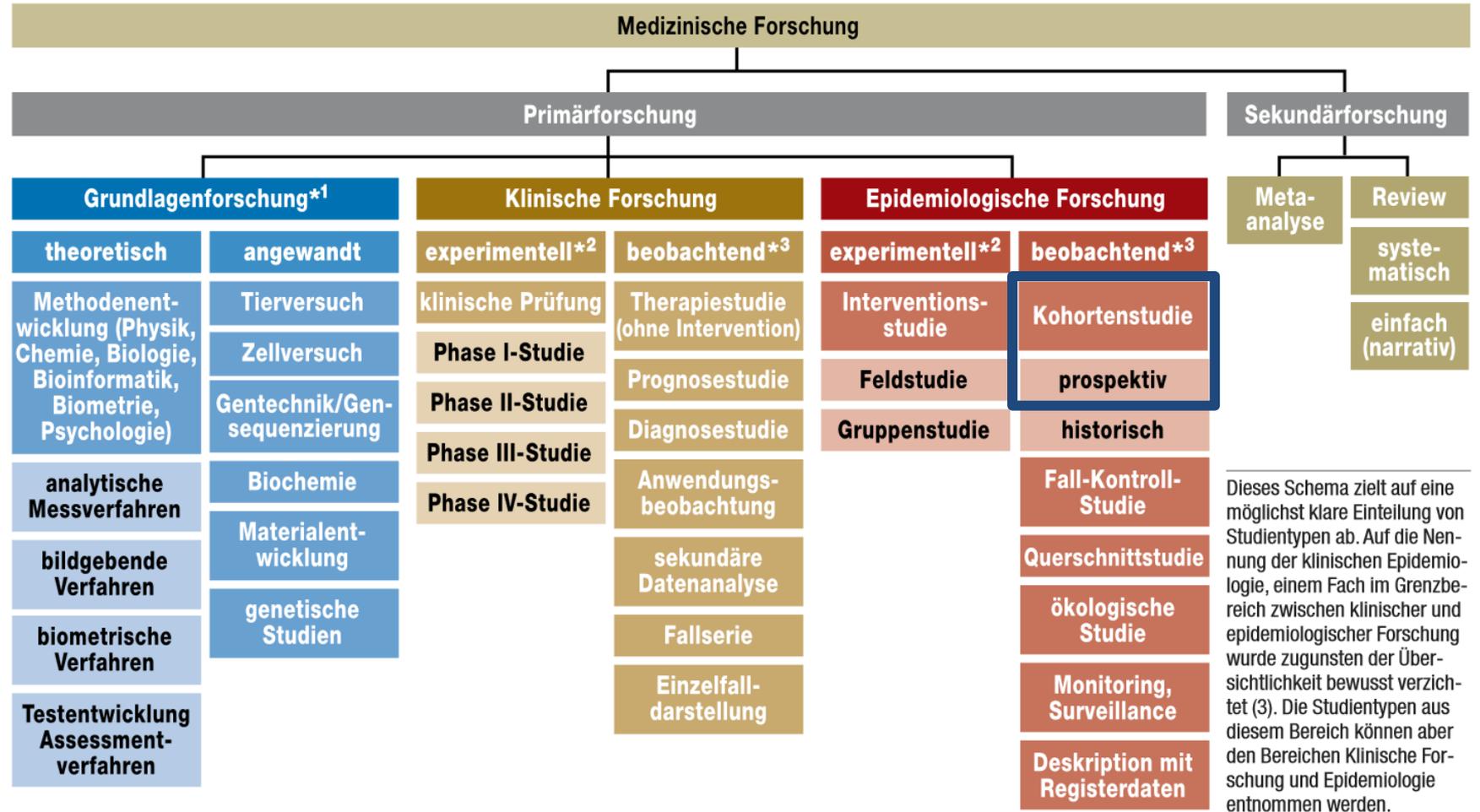
Funktionsbereich Qualität in der Gesundheitsforschung

1. **Datenqualitätsbewertungen sind kontextspezifisch**
2. Anforderungen einer epidemiologischen Kohortenstudie
3. Implementation von Datenqualitätsbewertungen
4. Ausblick und Fazit



Anwendungsszenarien

GRAFIK 1

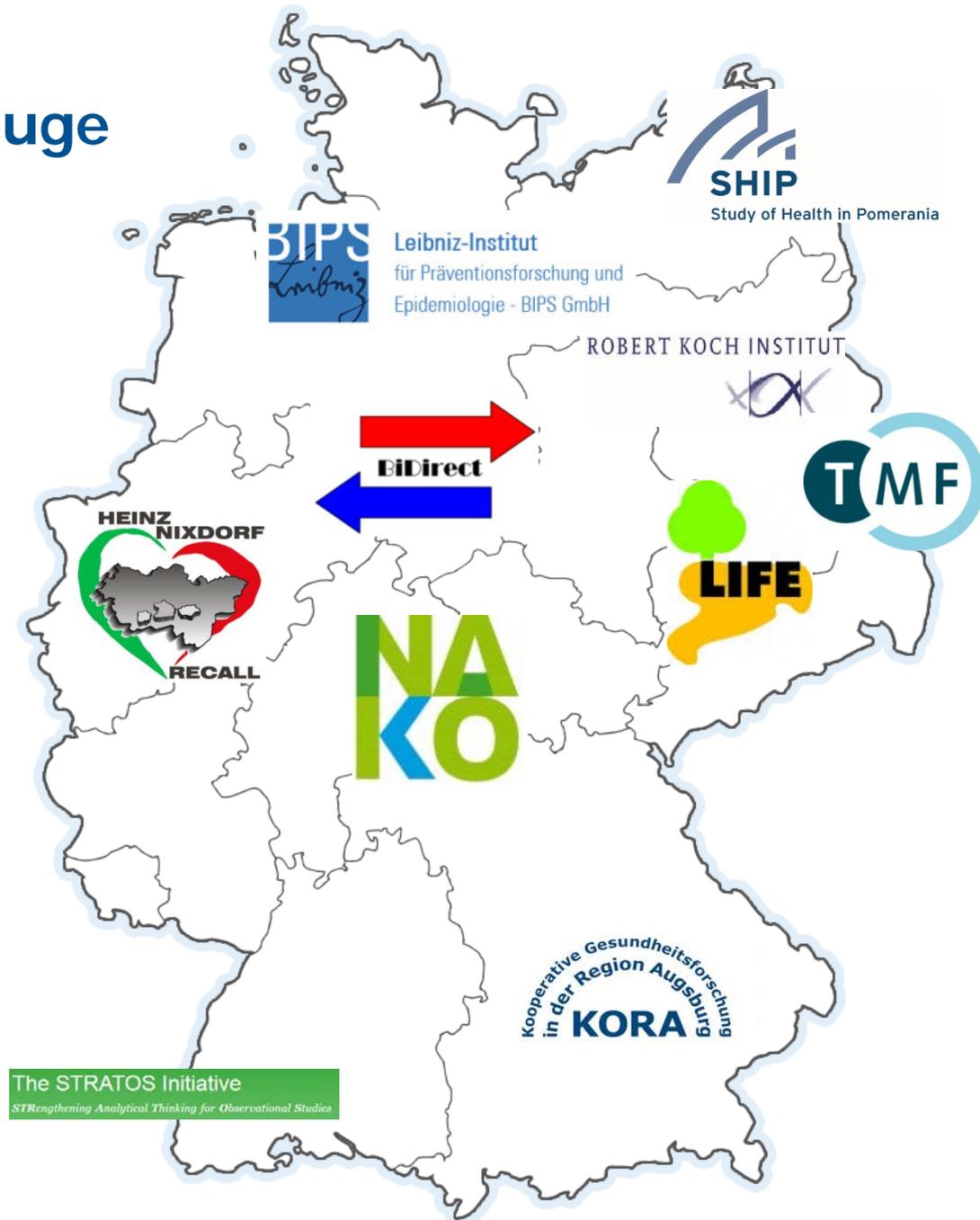


Einteilung verschiedener Studientypen

*¹ häufig synonym verwendet: Experimentelle Forschung; *² analoger Begriff: interventionell; *³ analoger Begriff: nicht interventionell/nicht experimentell

Standards und Werkzeuge zur Beurteilung der Datenqualität in komplexen epidemiologischen Studien

Schmidt, Carsten Oliver; Prof. Dr. Dr. Bamberg, Fabian; Prof. Dr. Berger, Klaus; Prof. Dr. Hoffmann, Wolfgang; Prof. Dr. Jöckel, Karl-Heinz; Prof. Dr. Kurth, Bärbel-Maria; Prof. Dr. Löffler, Markus; Prof. Dr. Meisinger, Christa; Prof. Dr. Pigeot, Iris; Prof. Dr. Stausberg, Jürgen; Prof. Dr.



Überarbeitung der TMF Indikatoren

Ziel: Etablierung von Standards und Werkzeugen zur Datenqualität für **Kohortenstudien**

Initial: Mapping und Überarbeitung bestehender TMF Indikatoren

Anforderungen

- vergleichbare Komplexität
- wenig Überlappung
- hierarchische Strukturierung



Dimensionen für Datenqualität in Kohortenstudien

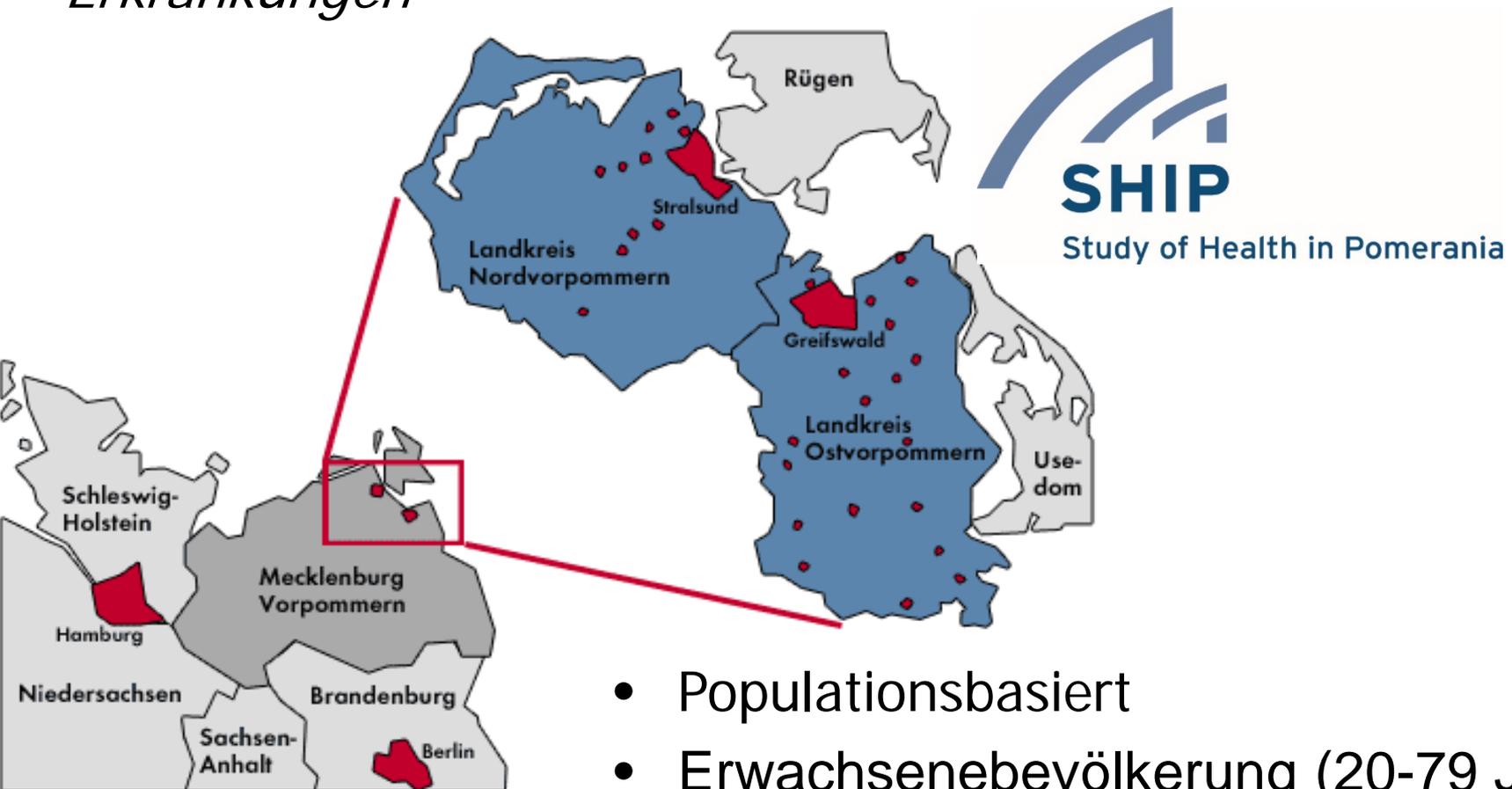
Beschreibung	Beispiele
Integrität	
Formelle Eignung für Qualitätsbewertungen	Studien-, Metadatenintegrität
Vollständigkeit	
Welche Daten sind verfügbar (Ausmaß, Einflussfaktoren)?	Unit / Segment / Item Missingness
Konsistenz	
Sind Daten frei von Widersprüchen und Konventionsbrüchen?	Widersprüchliche / unerlaubte Werte
Genauigkeit	
Stimmen Daten mit erwarteten Werten oder Verteilungen überein?	Untersucher- / Geräteeffekte, Zeittrends, Extremwerte

1. Datenqualitätsbewertungen sind kontextspezifisch
- 2. Anforderungen einer epidemiologischen Kohortenstudie**
3. Implementation von Datenqualitätsbewertungen
4. Ausblick und Fazit

Study of Health in Pomerania (SHIP) untersucht ...

- Prävalenz und Inzidenz von sowie
- Zusammenhänge zwischen

Risikofaktoren, subklinischen Auffälligkeiten und manifesten Erkrankungen



- Populationsbasiert
- Erwachsenebevölkerung (20-79 Jahre)

SHIP Kohorten und Untersuchungswellen

SHIP-0
Baseline

n= 4308

1997-2001

SHIP-1
5 J. FU

n= 3300

2002-2006

SHIP-2
10 J. FU

n= 2333

2008-2012

SHIP-3
15 J. FU

n= 1718

2014-2019

SHIP-4
20 J. FU

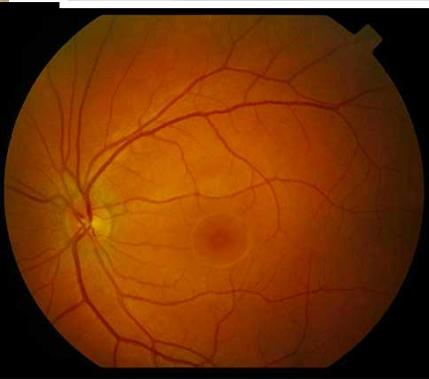
2019-2020

TREND-0
Baseline

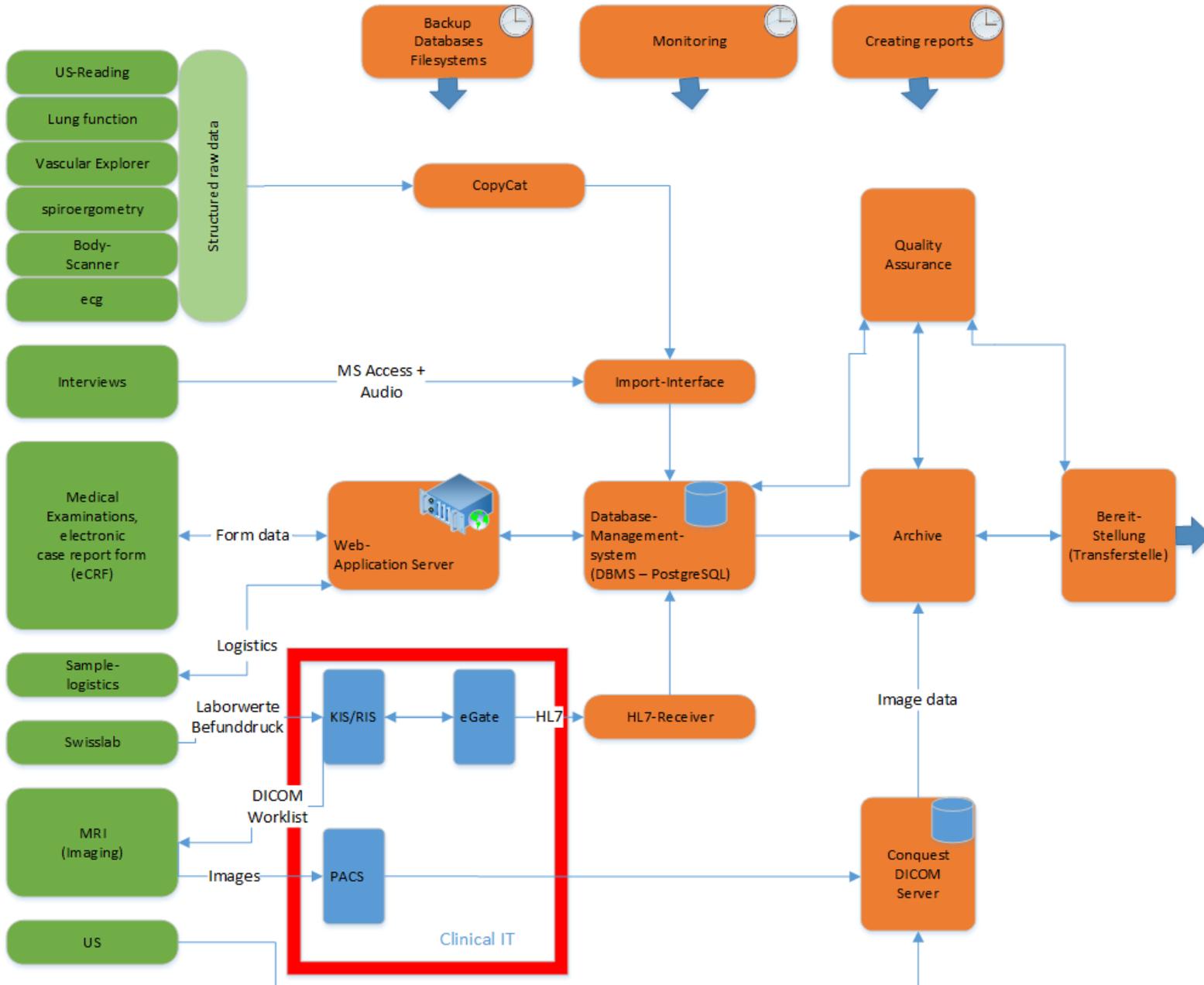
n= 4422

TREND-1
5 J. FU

n= 2507



Anforderungen IT und Datenmanagement



Anforderungen Qualitätsmanagement (Auszug)

Vor der Datenerhebung

Studienhandbücher (SOPs) und Checklisten

Untersucher: Schulung, (Re-)Zertifizierung

Geräte: Kalibrierung, -vergleich, -wartung

Pilotierung, Prätest

...

Während und nach der Datenerhebung

Kontrolle der Teilnahmevoraussetzungen

Standardisierte Dateneingabe

Mehrfachbefundungen (v.a. in Bildgebung): Intra- und Interreadervergleich

Site Visits durch Qualitätsverantwortlichen

Datenmanagement, syntaxbasierte Datenbereinigungen

Qualitätsberichte

...

1. Datenqualitätsbewertungen sind kontextspezifisch
2. Anforderungen einer epidemiologischen Kohortenstudie
- 3. Implementation von Datenqualitätsbewertungen**
4. Ausblick und Fazit



Datenqualitätsbewertung:

- Teilnahme
- Fehlende Werte
- Grenzüberschreitungen
- Extremwerte
- Untersuchereffekte
- Geräteeffekte
- Zeittrends

Vollständigkeit

Konsistenz, Genauigkeit

Genauigkeit

Dokumentation:

- Zertifizierungen
- Schulungen
- Datenqualitätsmaßnahmen
- ABSCHLUSS: Bewertung gemäß Ampelsystem und Qualitätshinweise

Hintergrundinformation

Entscheidungen

Framework: Initiale Datenanalyse

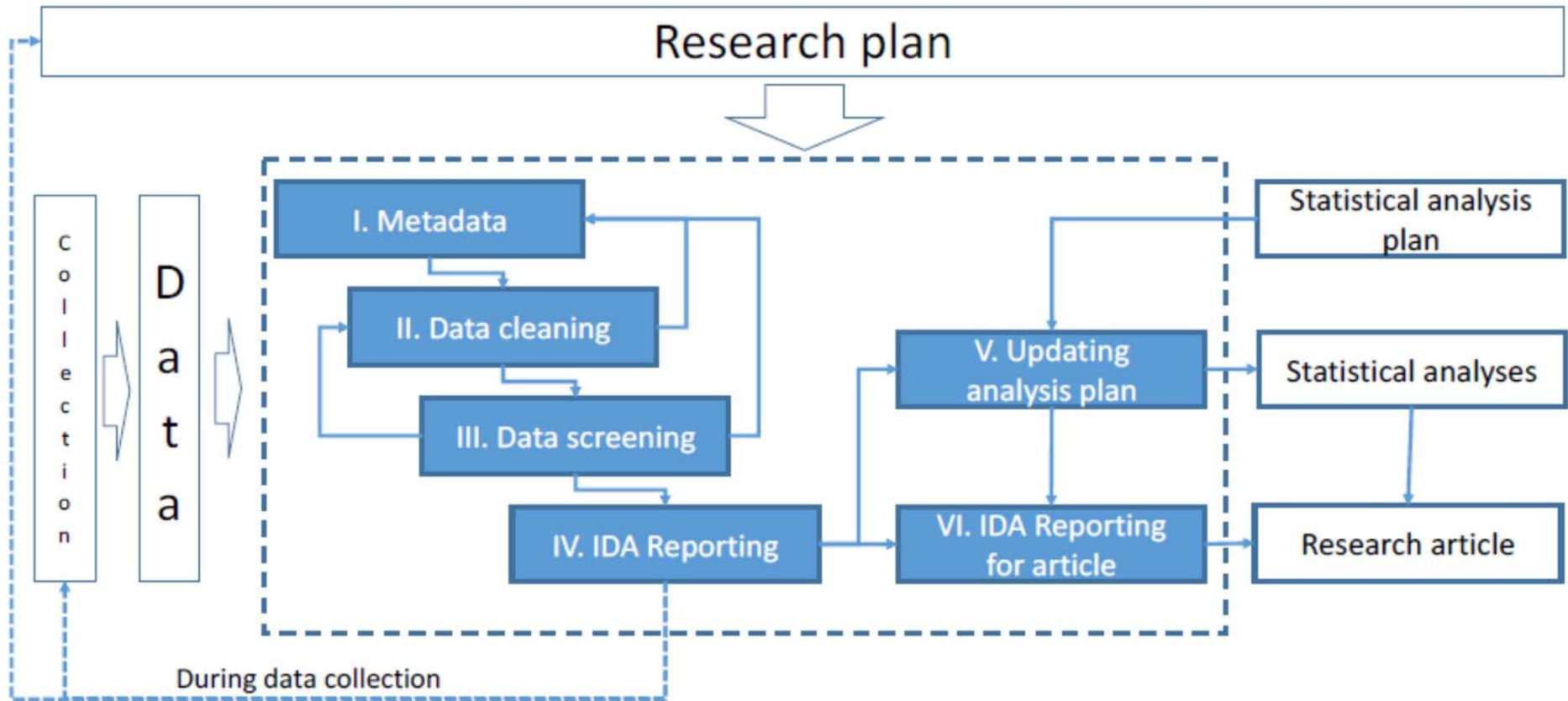
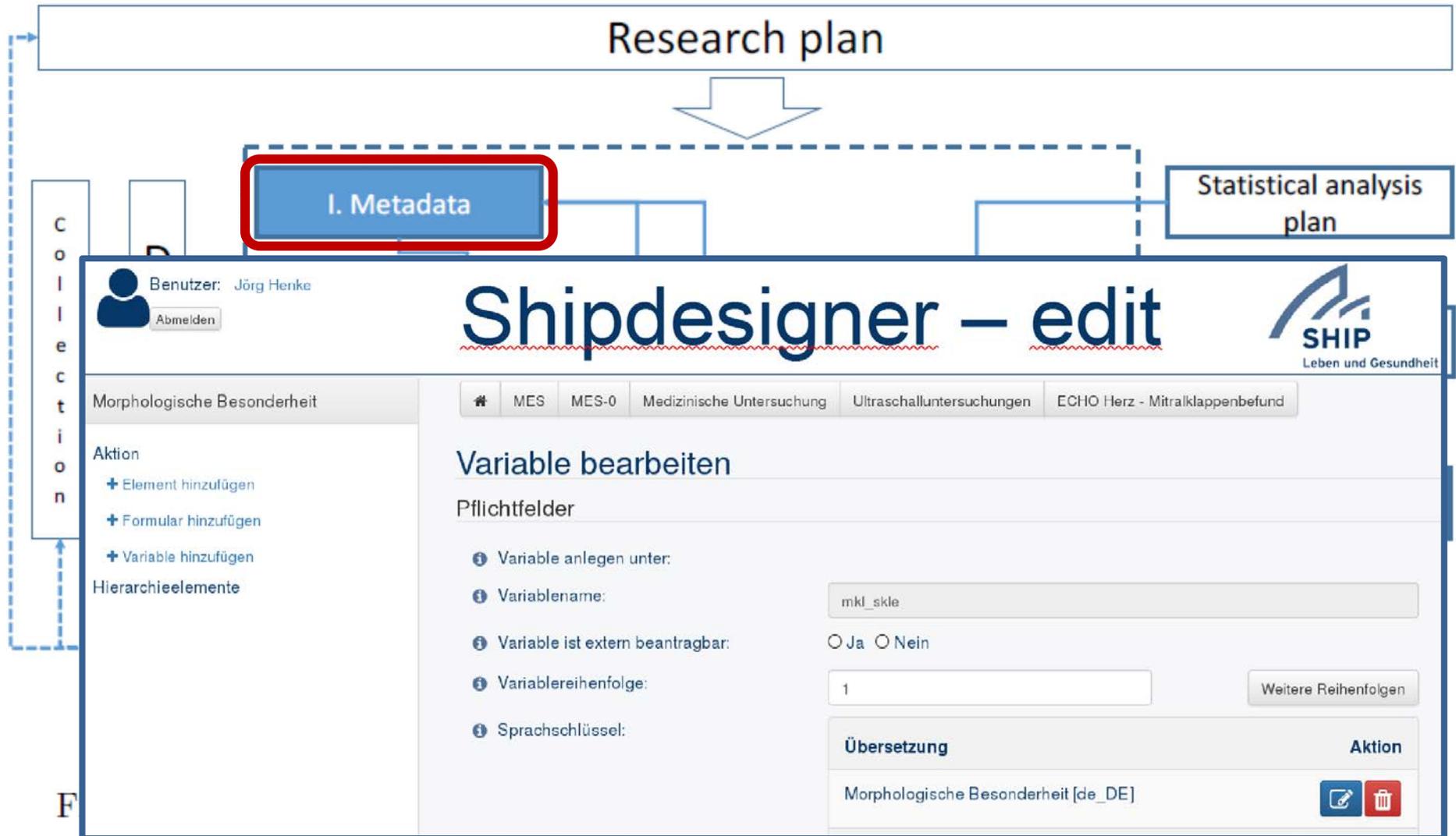


Figure 1: The main connections between the IDA steps and external components

Framework: Initiale Datenanalyse



Metadatenmanagement: Variablenattribute



Standard Operating Procedure (SOP)

Titel: Überarbeitung der Variablenattribute in SHIP zum Zwecke der automatischen Qualitätsberichterstattung in Square²

3	BEARBEITUNG EINZELNER VARIABLENATTRIBUTE	6
3.1	Variablenlabels	6
3.2	Datentyp	8
3.3	Plausibilitäts- und Zulässigkeitsgrenzen	8
3.4	Werteliste	9
3.5	Referenzkategorien	10
3.6	Variablenzuordnung	11
3.7	Missing- und Jumpcodes	12
3.8	Variablenreihenfolge	12

Framework: Initiale Datenanalyse

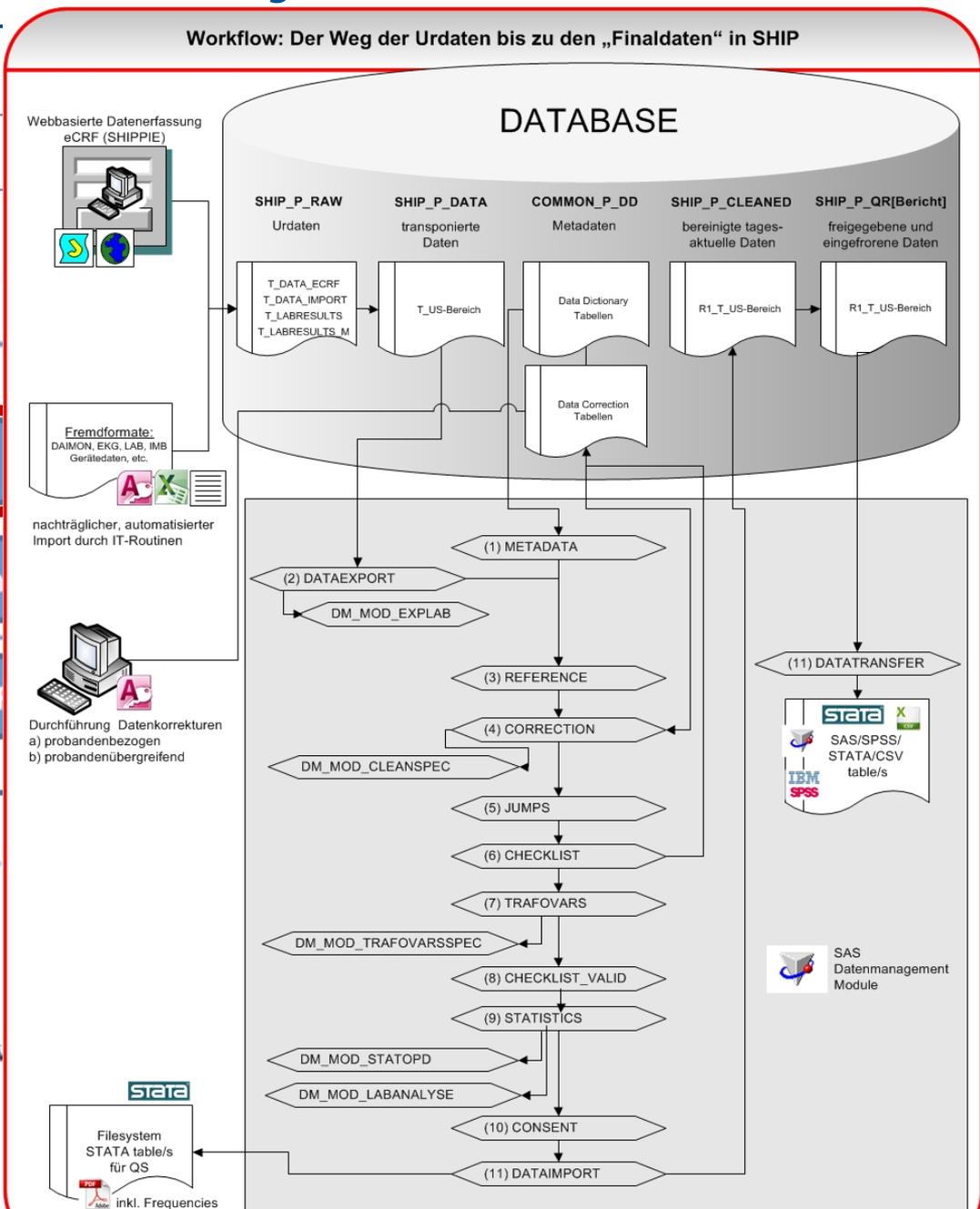
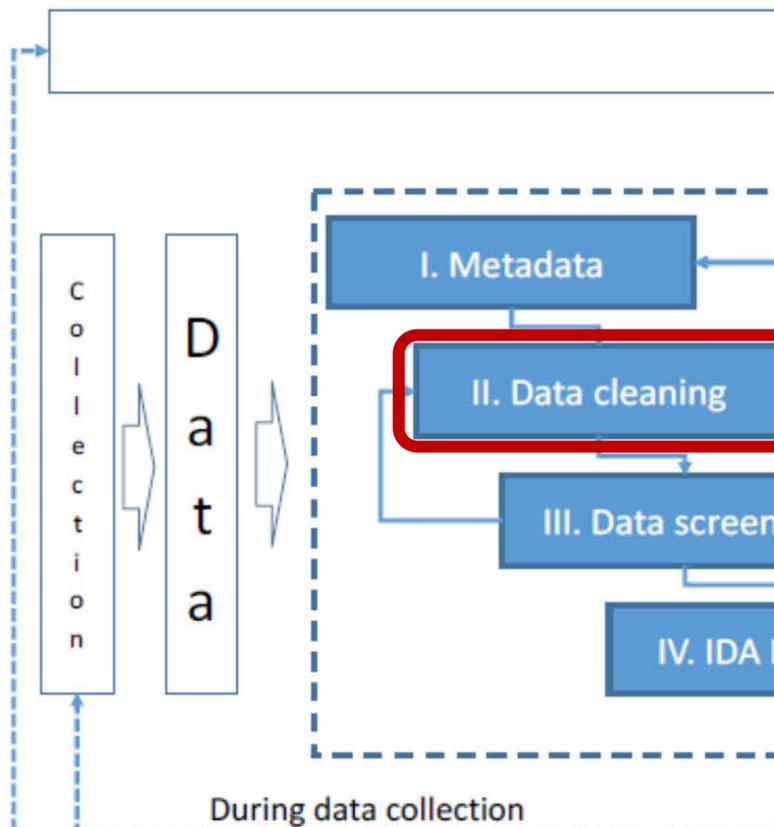


Figure 1: The main connections between

Framework: Initiale Datenanalyse

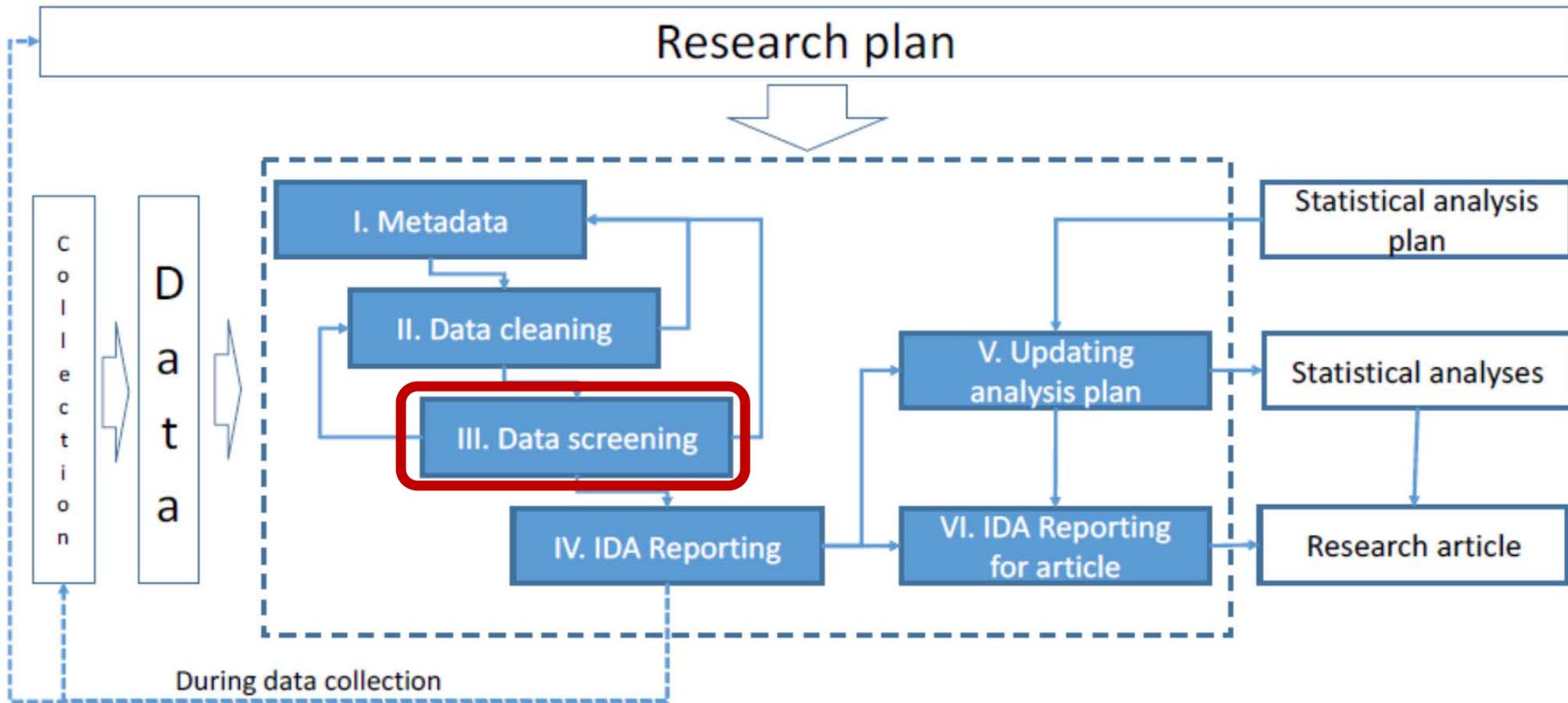


Figure 1: The main connections between the IDA steps and external components

Pipeline 1: Square² (Ship-QUALity Reports)

Detailliertes Rechte- und Rollenkonzept

Studienübersicht Studienstruktur Variablen-Gruppe Datenmanagement Statistik Vorlagen Qualitätsberichte Administration

Strukturierter Workflow

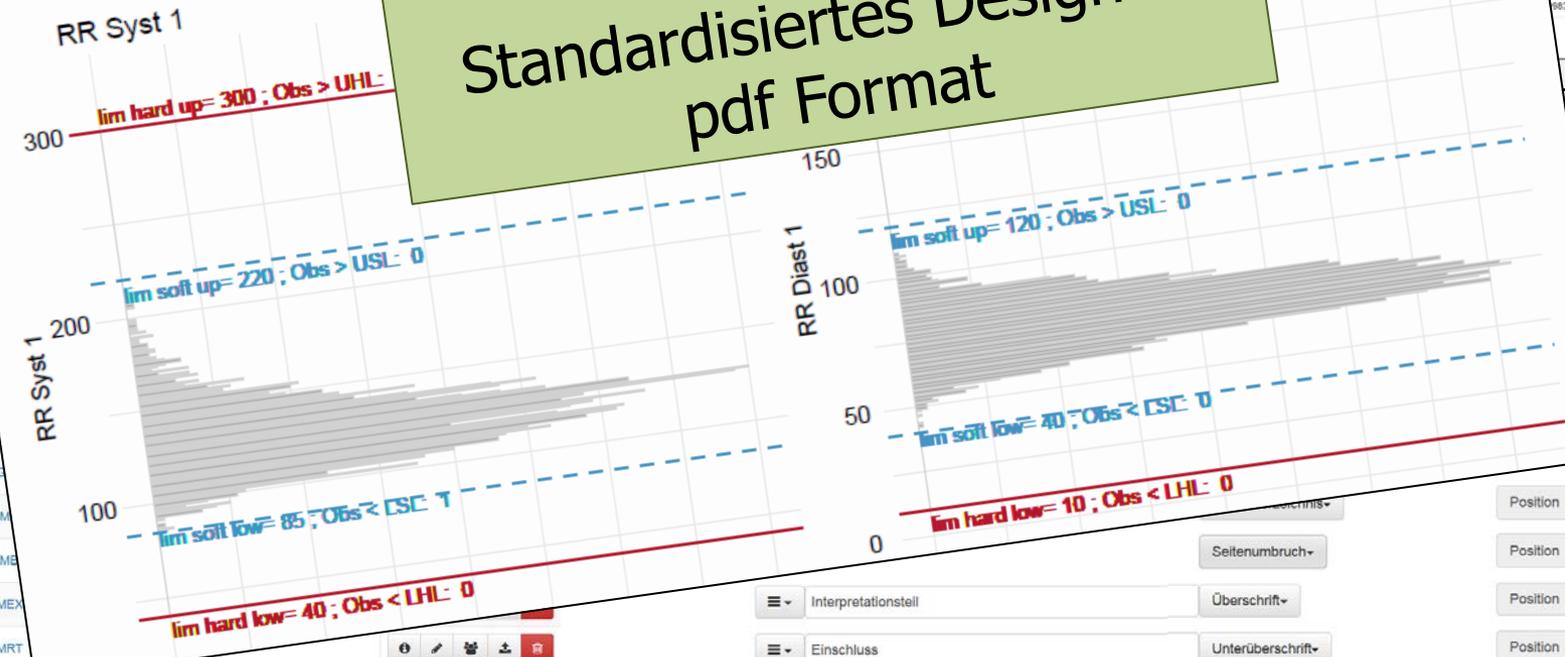
Square² - SHIP Quality Reports 2

Version 1.3.4.2 (PROD)
R version 3.3.3 (2017-03-06)
Square²

2.3 Verteilung und Plausibilitätsgrenzen (Berichtszeitraum)

Die Histogramme zeigen die Werteverteilung so wie die spezifizierten (Plausibilitäts- (blau) und Zulässigkeitsgrenzen (rot) während des Berichtszeitraums.

BERICHT:
Standardisiertes Design in pdf Format



St

Studien

- SHIP
- SHIP
- TREND
- MES
- TREND

Hierar

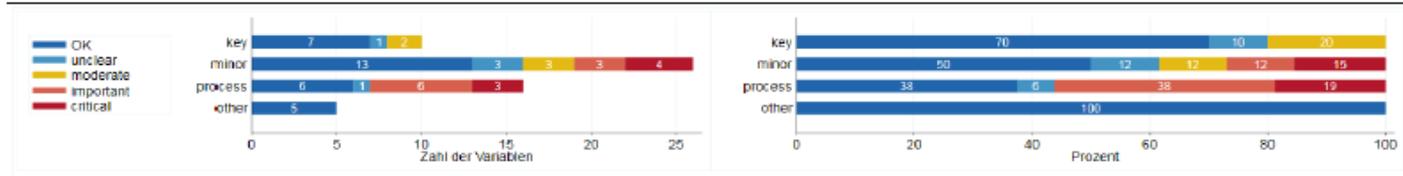
Reihenfolge	IT	Na
100		
270		
170		
160		
110		
190		

Interpretationsteil	Seitenumbruch	Überschrift	Position 2
Einschluss		Unterüberschrift	Position 3
		Kommentar / Interpretation der Ergebnisse	Position 4
			Position 5
			Position 7

Pipeline 2: Stata Reporter

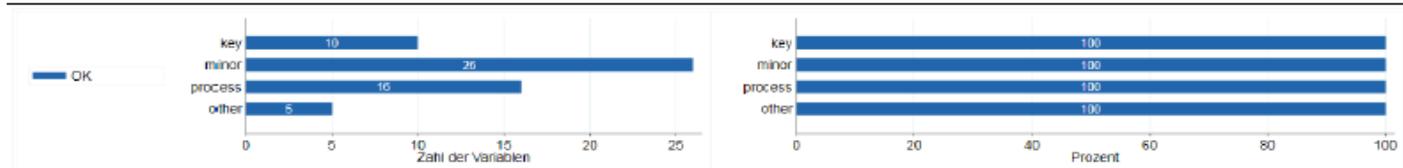
Überblick zu Datenqualität

Variablen mit Datenqualitäts Issues



Dieser Graph zeigt die höchste Problemkategorie je Variable an.

Fehlende Werte



Categories for Missing %: OK : <1; moderate : <10; important : <25; critical : >=25;

Unplausible Werte



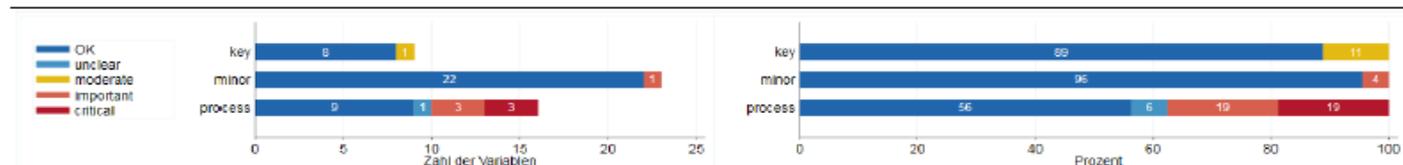
Categories for Implausible values %: OK : =0; unclear : <2; moderate : <5; important : <10; critical : >=10;

Unerwartete Effekte durch Untersucher



Categories for Fixed effects ICC: OK : <0.02; unclear : <.03; moderate : <0.05; important : <.1; critical : >=.1;

Unerwartete Effekte durch Gerät



Categories for Fixed effects ICC: OK : <0.02; unclear : <.03; moderate : <0.05; important : <.1; critical : >=.1;

Detailgrad abhängig von Typ der Variable

Cluster und Zeiteffekte für: Untersucher ech_usnr

Primäre Variablen	ICC	Cluster verwendet	Cluster alle	Kleinstes Cluster	Mittelwert Cluster	Größter Cluster	N in Analyse
aki_ins m Echo Aortenkl. Insuffizienz JN modified	<.01	4	4	326	626	1034	2505
aki_skle m Echo Aortenkl. Sklerose JN modified	.02	4	4	326	626	1034	2505
ech_aki m Echo Herz Aortenklappenbefund JN modified	.01	4	4	326	626	1034	2505
ech_mki m Echo Herz Mitralklappenbefund JN modified	.01	4	4	327	626	1034	2506
ech_ik m Echo Herz Trikuspidalklappeninsuffizienz JN modified	.01	4	4	326	626	1035	2506
mki_ins m Echo Mitralkl. Insuffizienz JN modified	.01	4	4	327	626	1034	2506
mki_skle m Echo Mitralkl. Sklerose JN modified	.04	4	4	327	626	1034	2506



Tabellarische Ergebnisse für alle Variablen

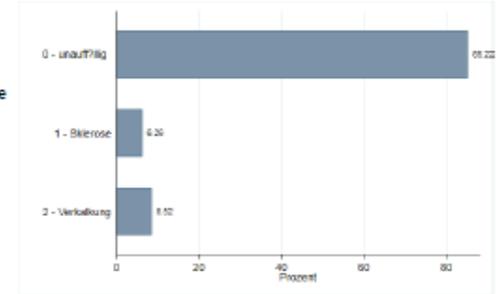
Detaillierte Übersicht für primäre und Prozessvariablen



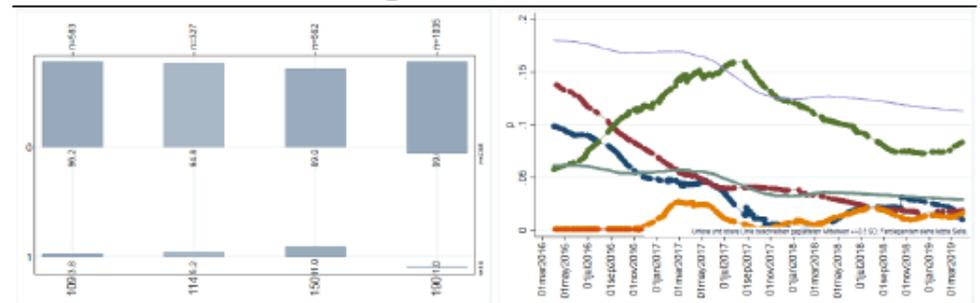
Ergebnisse für Variable: mkl_skle

Primäre Variable mit Label: Echo Mitralkl Sklerose JN

Maße	Ausgangsvariable	Modifizierte Variable
Observations	751	2506
Missings	1	1
		95.57 % Reference
		4.43 % Event

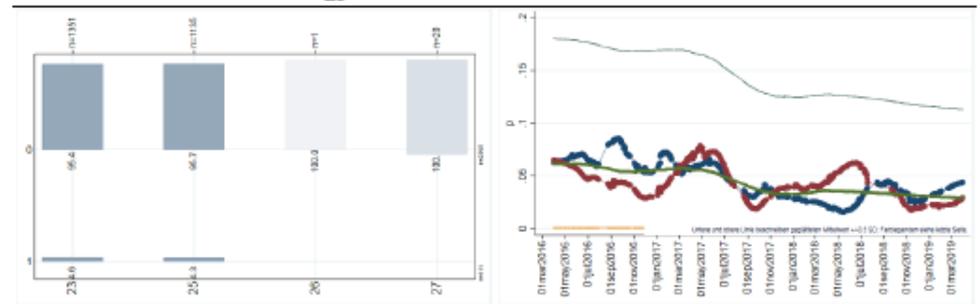


Cluster und Zeiteffekte für: Untersucher ech_usnr



Automatische Umkodierungen: Missings umkodiert (n=1755); Kategorie geändert (n=64); Referenzkategorie der Ausgangsvariable ist: 0

Cluster und Zeiteffekte für: Gerät ech_grid



Automatische Umkodierungen: Missings umkodiert (n=1755); Kategorie geändert (n=64); Referenzkategorie der Ausgangsvariable ist: 0

1. Datenqualitätsbewertungen sind kontextspezifisch
2. Anforderungen einer epidemiologischen Kohortenstudie am Beispiel SHIP
3. Implementation am Beispiel von SHIP
4. **Ausblick und Fazit**



Fazit

1. Anwendungskontext ist wichtig
2. Dimensionen für Indikatoren bei Kohortenstudien:
 - a) Integrität
 - b) Vollständigkeit
 - c) Konsistenz
 - d) Genauigkeit
3. Wichtige Aspekte:
 - a) Solides Metadatenmanagement
 - b) Effektive IT-Infrastruktur
 - c) Einheitliche Konzepte zur Datenqualitätsanalyse
 - d) Standards für Routinen, Berichte und Tools
 - e) Automatisierung
4. Ziele: Steigerung der Datenqualität, Transparenz und Vergleichbarkeit



<http://www.medizin.uni-greifswald.de/icm/>

birgit.schauer@uni-greifswald.de; carsten.schmidt@uni-greifswald.de