

# Data-Management



Im Institut für Medizinische Informatik,  
Biometrie und Epidemiologie

Claudia Ose

# Anforderungen an unsere SOPs

- GCP-gemäß
- Anwendbar für
  - Zulassungsstudien
  - Anwendungsbeobachtungen
  - Universitäre Studien
  - Epidemiologische Studien
- Preiswert
- Umsetzbar
- Praxisnah

# Praktische Umsetzung

## ***Erstellen eines studienspezifischen Data-Management-Plans***

- In welchem Verzeichnis sind die Daten gespeichert und wer erhält Zugang zu den Daten?
- Datenbanksystem
- Wer führt die Sichtkontrolle durch?
- Art der Dateneingabe
- Dateneingabeanleitung (Erst-Entwurfs)
- Datenerfasser
- Wer entscheidet bei Problemen in der Datenerfassung?
- Welche Codiersysteme?
- Notwendigkeit und Umfang eines Mahnsystems
- Unterstützung des Monitors durch den Data-Manager
- Wieviel Prozent der Prüfbögen werden stichprobenartig mit der Datenbank verglichen?

# Art der Dateneingabe

- Einfach
- Doppelt
- Key-to-Key verification
- File-to-File matching
  
- Womit
  - ACCESS
  - SAS

# Plausibilitätsprüfung in der DB

- Erlaubte Werte/Wertebereiche sind direkt in die Datenbank einzubauen
- Missings müssen unterscheidbar sein zwischen dem Fall “(noch) fehlender Wert = Null” und dem Fall “Wert nicht ermittelbar/Test nicht durchführbar = -9
- Bei einfacher Dateneingabe sind die Plausibilitätsprüfungen zu erweitern

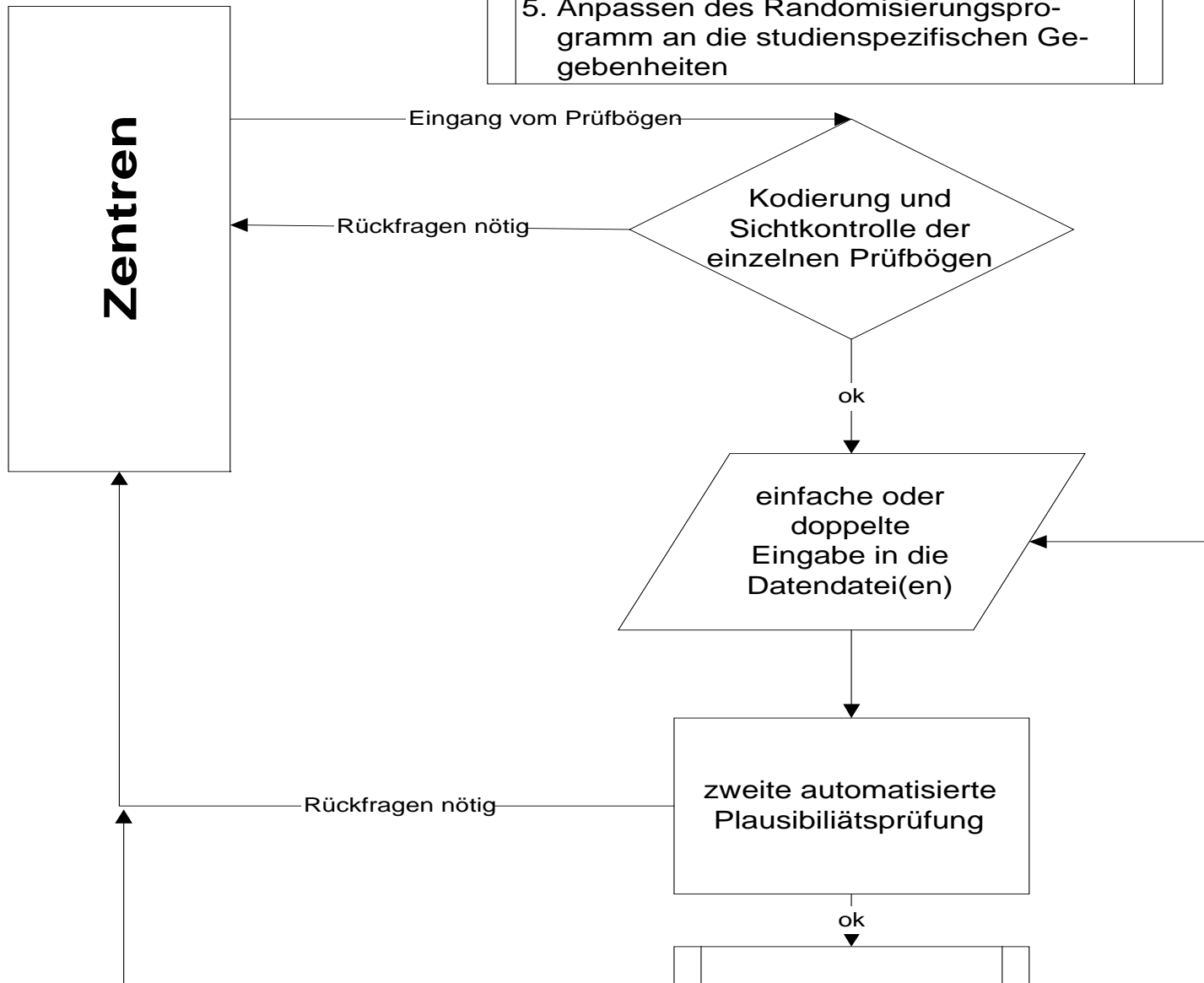
# Verzeichnisstrukturen

- daten (Datendateien)
  - erst (Datendateien für die Ersterfassung)
  - zweit (Datendateien für die Zweiterfassung)
  - kontr1 (Datendateien für die Plausibilitätsprüfungen)
    - Ggf. kontr2 ...
  - arbeit (Datendateien nach vollständiger Durchführung der Plausibilitätsprüfungen), schreibgeschützt

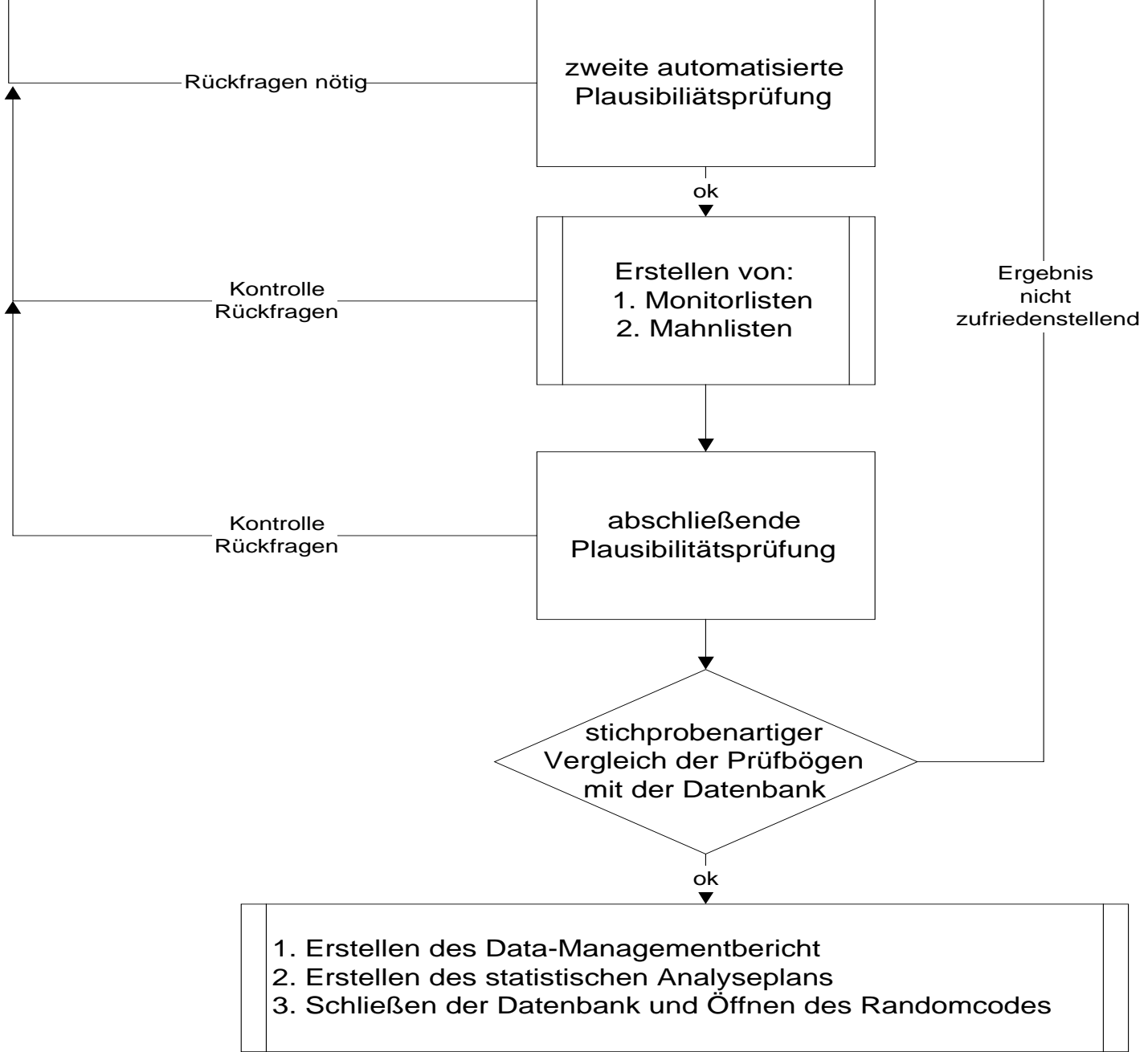
# Variablennamen

- Variablen, die zu einem Komplex gehören erhalten einen gemeinsamen Namen und unterscheiden sich nur durch die Endziffer (z.B. rrsyst1, rrsyst2, rrsyst3).
- Datumsfelder enden mit dem Zeichenfolge “\_dat” (z.B. geb\_dat, ran\_dat).
  - Jahr 4-stellig
  - Monat in der Datumsausgabe im Format “Jan/Feb/Mar...”
- Freitext und der dazugehörige Code erhalten einen gemeinsamen Namen. Durch die Endung “\_FT” und “\_CD” wird der Freitext bzw. der Code identifiziert (z.B. Begleiterkrankung: BE\_FT und BE\_CD).
- Variablen, die programmtechnisch modifiziert werden, behalten den selben Namen (ggf. verkürzt), nur die Endung beschreibt die Transformation z.B.
- “\_lg” logarithmiert zur Potenz 10
- “\_ln” natürlicher Logarithmus
- “\_5K” klassiert in 5er Klassen
- “\_10K” klassiert in 10er Klassen
- Die Variable darf nicht mit einer Ziffer beginnen. Als Sonderzeichen darf nur der Underscore “\_” benutzt werden.

1. Erstellen des Data-Managementplans
2. Art der Dateneingabe
3. Auswahl der Codiersysteme
4. Aufbau der Datenbank
5. Anpassen des Randomisierungsprogramm an die studienspezifischen Gegebenheiten







# Datenerfassung

- Erfassung aller Bögen im Tracking-System
- Erfassung der Daten durch studentische Hilfskräfte der Medizin
- (pro Hilfskraft) Eingabe der ersten Bögen gemeinsam, danach
- Selbstständige Eingabe, danach
  - Kontrolle durch den Data-Manager (wird dok.)
  - Bei vielen Fehlern: Kontrollen verstärken
- In jedem Fall Querykontrolle und Weiterschreibung der Eingabeanleitung

# Kommunikation mit dem Monitor

- Bei Prüfplanverletzungen / Verstöße gegen die Ein-/Ausschlusskriterien / SUEs
  - Information des Monitors
  - Information des Prüfarztes
  - Dok. der Maßnahmen zur Fehlerbeseitigung
- zentrumsspezifische Listen für den Monitor zum Mitnehmen
  - Kontrolle von Streuungen, Auffälligkeiten im Zentrum
  - Queries
  - Anzahl der UEs/Begleiterkrankungen

# Abschließende Plausibilitätskontrolle

- Gibt es noch offene Fragen / Implausibilitäten?
- Sind alle fehlenden Werte ergänzt oder alle fehlenden Werte geklärt?
- Stehen noch Bögen aus?
- Sind alle abnormen Laborwerte als solche beschrieben und bewertet?
- Sind alle UEs als solche beschrieben (versteckte UEs in Bemerkungen, Änderungen der Begleitmedikationen, Abbrüchen)?
- Sind alle schwerwiegenden UEs vollständig und korrekt beschrieben?
- Sind alle Studienabbrüche als solche beschrieben?
- Sind alle Protokollverletzungen benannt?
- Stimmt der vergebene Code mit den Freitextangaben überein?

# Transformationen

- Jede Transformation muss überprüft werden z.B.
  - Transfer nach SAS
  - Altersklassen bilden
  - Logarithmieren
  - Formatzuweisung

# Audit-Trail

- Separate Erfassung von Erst- und Zweiteingabe
  - Entscheidungsprobleme werden auf einer separaten Liste dok.
  - Falls key-to-key verification → autom. elektr. Audit trail in ACCESS
  - Falls File-to-File matching → Schriftliche Dok. der Änderungen
- Danach kopieren der Datei in das „Kontr N“-Verzeichnis
  - Alle Änderungen müssen nachvollziehbar und dokumentiert sein.
  - Direkte Änderungen in der Datendatei ohne zusätzliche Dokumentation dürfen nur vorgenommen werden, wenn die Korrektur im Prüfbogen vom Prüfarzt dokumentiert wurde.
- Bei der Erfassung mit SAS sollte ein „Audit Trail“ mitlaufen

# Data-Management-Bericht

- Formate und Dateistruktur
  - Datenbankstruktur
  - Variablenbezeichnungen, Formate und Codes.
  - Prüfbogenmuster mit allen Variablen und den dazugehörigen Formaten beschriftet
  - Beschreibung aller Programme im Rahmen des Data-Managements (incl. Log`s und List`s)
  - Laboreinheiten und Normbereiche
- Ergebnisse von Data Management-Kontrollen
  - Ausdruck der Tracking-Datei, patientenbezogen
  - Auflistung der Datendiskrepanzen zwischen Erst- und Zweiterfassung und nach Plausibilitätskontrollen
  - Fehler nach Datenüberprüfung
  - Dokumentation der Überprüfung der Datenerfassung
  - Transformationsprogramm(e)
  - Query-Listing
  - Auflistung, welche Rückfragen von den Zentren nicht beantwortet worden sind
- Datenprobleme
  - Auflistung und ggf. Begründung ungelöster Datenprobleme
  - Protokollverletzungen
  - Auflistung der fehlenden Daten (variablenbezogen)
  - Besonderheiten von Zentren

# Danach

- Festlegung der Kollektive
  - Verträglichkeit, ITT, PP
- Schließen der Datenbank
- Erstellen / Validierung des statistischen Analyseplans
- Eingabe des Randomcodes
- Auswertung
- Bericht