

Erstellung einer Integrierten Datenbank mit SAS für die gepoolte Analyse mehrere klinischer Studien

Thomas Bruckner, Monika Engemann, Doris Kolb, Stefan Pautz

Abteilung Arzneimittelsicherheit
Pharmaceutical Research Associates GmbH
Telefon: 0621 / 87 82 – 0
www.prainternational.com
eMail: BrucknerThomas@PRAIntl.com

Abstract

Während der klinischen Entwicklung eines Medikamentes werden mehrere Studien mit unterschiedlichen Patientenzahlen und Fragestellungen durchgeführt. Die Daten werden in studienspezifischen Datenbanken erfaßt. Im nachfolgenden Beitrag wird die Erstellung einer Integrierten Datenbank mittels SAS aufgezeigt und die Lösung möglicher Probleme beschrieben. Diese Integrierte Datenbank ist für den Zulassungsprozeß von großer Bedeutung.

Einleitung

Bevor ein Medikament zugelassen werden kann, müssen eine Reihe von Klinischen Studien (Phase I – III) zu diesem Präparat durchgeführt werden. Diese Studien können zwischen 20 und mehreren tausend Patienten umfassen. Die Daten zu diesen Studien liegen zum Teil in sehr unterschiedlichen Datenstrukturen vor. Eine Zusammenfassung dieser Daten für Auswertungszwecke ist in diesem Falle nicht so ohne weiteres möglich. Das Standardisieren dieser Daten ist jedoch aus folgenden Gründen sinnvoll:

- Analyse seltener Unerwünschter Ereignisse
- Erhöhung der Präzision des geschätzten Behandlungseffektes
- Erlangung des notwendigen Stichprobenumfangs für Subgruppenanalysen
- Beantwortung neuer Fragen, die bisher in den existierenden Studien nicht betrachtet wurden oder für die in den einzelnen, schon existierenden Studien die Fallzahl nicht ausreicht.
- Erreichung einer größeren Unabhängigkeit der Ergebnisse von speziellen Studiensituationen (wie z.B. Prüferärzte, untersuchte Patientenpopulationen etc.)
- Ad-hoc Auswertung für die Beantwortung von Rückfragen (z.B. von Zulassungsbehörden)

Ein eindrucksvolles Beispiel für die schnell sehr groß werdenden Fallzahlen ist in nachfolgender Tabelle aufgeführt. Diese enthält die Anzahl von Beobachtungen, die notwendig sind, um mindestens ein Ereignis zu entdecken.

Güte (Power)	Wahrscheinlichkeit von Ereignissen			
	1 %	0,1 %	0,01 %	0,001 %
95 %	300	2 995	29 956	299 572
80 %	161	1 608	16 094	160 043
50 %	69	693	6 932	69 315

Bei der Auswertung einer einzelnen Studie besteht daher die Gefahr, seltene Ereignisse nicht zu erkennen. Manche seltenen Ereignisse treten z.B. nur 1x in 5 000 Patienten auf, d.h. diese Ereignisse kommen meist in kleineren Populationen nicht vor. Nicht zuletzt deshalb finden meta-analytische Methoden immer öfter Anwendung bei der Bewertung von Arzneimittelwirkung und -risiken. Für diese Auswertung ist es notwendig, eine einheitlich definierte Datenbasis zur Verfügung zu haben. Im folgenden beschreiben wir Ausgangssituation und Vorgehensweise bei der Erstellung einer Integrierten Datenbank mittels SAS.

Ausgangssituation

Beim Zusammenführen der einzelnen Studien können verschiedene Probleme auftreten, die eine Konvertierung der einzelnen Studien in eine standardisierte Struktur sinnvoll erscheinen lassen:

Unterschiedliches Studiendesign

- Unterschiedliche Fragestellungen der Studien
- Unterschiedliche Vergleichsmedikation (Placebo, Standardmedikation)
- Unterschiedliche Dosierungen und / oder Applikationsformen (oral, i.m., i.v.)

Unterschiedliche Realisierung der Daten

- Daten liegen in unterschiedlichen Formaten vor (ASCII, DBASE, EXCEL, ORACLE, SAS etc.)
- Art und Anzahl der Dateien pro Studie sind unterschiedlich
- Die Datenstruktur ist bei gleicher Fragestellung unterschiedlich gelöst
- Variablen können bei gleicher Bedeutung verschiedene Realisierungen aufweisen (numerisch / alphanumerisch)

Kodierung

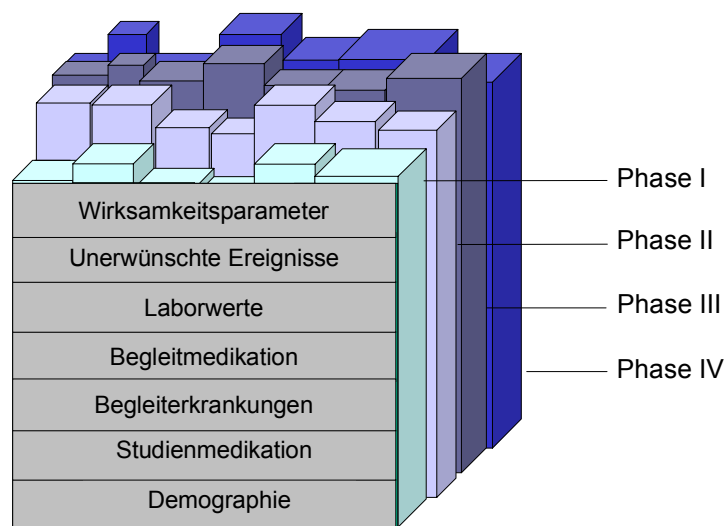
- Die zugrundeliegenden Kodierungs-Systeme sind verschieden (ICD9, ICD10, MedDRA, HARTS)
- Inkonsistenzen verursacht durch unterschiedliche Kodierungsprinzipien (“Splitting” / “Lumping”)

Verschiedene Maßeinheiten

- Der gleiche Parameter wird in verschiedenen Einheiten gemessen (z.B. Calcium in mg/dl oder mmol/l; Gewicht in lbs, kg)
- Verschiedene Referenzbereiche für den gleichen Laborparameter in unterschiedlichen Studien

Lösung

Folgende Abbildung stellt die verschiedenen Klinischen Phasen und die gemeinsamen Variablen symbolisch dar. Die Blöcke symbolisieren die einzelnen Studien. Die unterschiedlichen Grautöne stehen für die unterschiedlichen Phasen der Klinischen Entwicklung. Die überstehenden Blockenden symbolisieren die Variablen, die nicht standardisiert werden. Für die Integrierte Datenbank selektiert man die Variablen, die für die Auswertung des entsprechenden Medikaments von Interesse und Bedeutung sind. Für diese Daten wird eine einheitliche Datenstruktur definiert.



Bei der Erstellung der Integrierten Datenbank geht man folgendermaßen vor:

- Definition einer Integrierten Datenbankstruktur auf Basis der einzelnen Studien
- Erstellung von SAS-Programmen zur Konvertierung jeder Studie in diese neue Datenbankstruktur
- Definition von "Missingcodes", die deutlich machen, ob z.B. eine Variable in der zugrunde liegenden Studie nicht abgefragt wurde
- Vereinheitlichung der Variablenformate (z.B. alle Patientennummern alphanumerisch konvertieren, 0=nein 1=ja)

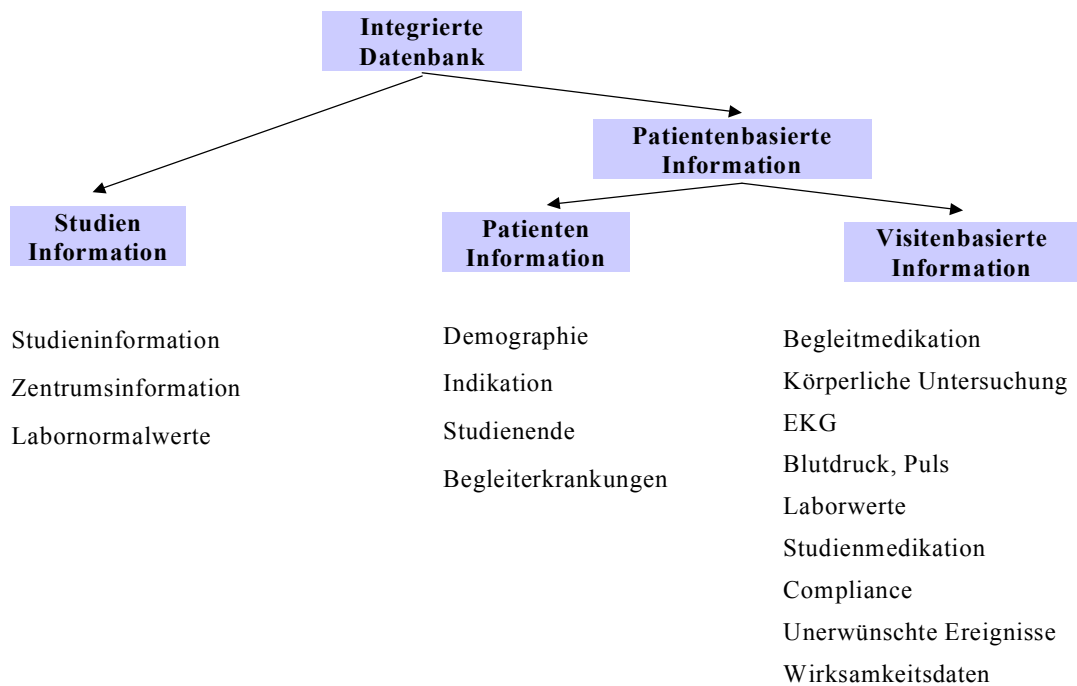
- Umwandlung von Maßeinheiten in standardisierte Einheiten
- Festlegung auf ein einheitliches Kodierungs-System und eventuell Umkodierung der Studien, die mit anderen Systemen kodiert wurden
- Überprüfung der Kodierungskonsistenz der gesamten Integrierten Datenbank
- Standardisierte Überprüfung der Plausibilität über alle Daten
- Erstellung von Datensätzen, die studienspezifische Informationen enthalten (z.B. Zentrumsinformationen, Labornormalwerte, Informationen zum Studiendesign)

Vorteile von SAS für die Erstellung einer Integrierten Datenbank

Für die Erstellung einer standardisierten Datenbank ist SAS aus folgenden Gründen geeignet:

- SAS ist plattformunabhängig
- Datenkonvertierung erfolgt mittels SAS Programmen und SAS Macros, d.h. es ist kein zusätzlicher Aufwand erforderlich, um die Daten in ein anderes Datenbanksystem zu überführen.
- Definition von Missingcodes ist möglich (z.B. ".N" für nicht zutreffend)
- Es ist kein Datentransfer für die statistische Auswertung nötig
- SAS ist ein validiertes System
- SAS Dateien werden von Zulassungsbehörden (z.B. FDA) akzeptiert
- Integration von Datenbankstruktur und Programmierumgebung

Struktur einer Integrierten Datenbank



Beispiel einer von PRA erstellten Integrierten Datenbank

Um den Umfang einer Integrierten Datenbank zu veranschaulichen, ist hier ein konkretes Beispiel angeführt:

Anzahl der Studien:	39
Anzahl der Patienten:	5 935
Anzahl der auswertbaren Patienten	4 498
Anzahl der Patienten in Phase I Studien:	351
Anzahl der Patienten in Phase II-III Studien:	4 147
Anzahl der unerwünschten Ereignisse:	7 124
Anzahl der Laborbeobachtungen:	762 645
Anzahl der Begleiterkrankungen:	15 322
Anzahl der Begleitmedikationen:	12 818

Diese Integrierte Datenbank war die Basis für Integrated Summaries of Safety and Efficacy. Auf Basis der Integrierten Datenbank konnten die Rückfragen der Zulassungsbehörden schnell beantwortet werden. Das Medikament wurde erfolgreich zugelassen.

Schlußbemerkung

Die Erstellung einer Integrierten Datenbank nach Beendigung der Klinischen Studien hat folgende Nachteile:

- Aufwendig in Bezug auf Programmierung
- Möglicher Zeitverlust im Zulassungsprozeß
- Konventionen müssen im nachhinein definiert werden

Die Planung einer Integrierten Datenbank vor der Durchführung Klinischer Studien bietet folgende Vorteile:

- Integrierte Datenbankstruktur ist die Basis für die Datenstruktur und CRF-Design der einzelnen Studien
- Standardisierte Datenerhebung
- Geringer Konvertierungsaufwand
- Ad-hoc Auswertungen sind während der Klinischen Prüfung des Präparates möglich
- Auswertungsprogramme können standardisiert werden
- Kodierungsprinzipien werden einheitlich für alle Studien definiert

Da die Zeit bis zur Zulassung eines Medikaments möglichst kurz sein sollte, gehen immer mehr Firmen dazu über, Standards im voraus zu definieren. Diese Standards umfassen z.B. CRF-Design, Integrierte Datenbankstruktur, studienspezifische Datenbankstruktur oder Kodierungsprinzipien. Diese Standards erleichtern die Erstellung einer Integrierten Datenbank. Die Integrierte Datenbank kann bereits während der Durchführung der Klinischen Studien erstellt werden. Diese Datenbank steht somit nach Beendigung der letzten zulassungsrelevanten Studie für studienübergreifende Auswertungen (Meta-Analysen) zur Verfügung.