

Metadatengetriebene statistische Auswertung klinischer Analysedaten

Florian Storch
HMS Analytical Software GmbH
Otto-Volger-Str. 3c
65843 Sulzbach
Florian.storch@analytical-software.de

Zusammenfassung

Der vor Ihnen liegende Beitrag beschäftigt mit sich mit dem Thema „Metadatengetriebene statistische Auswertung klinischer Analysedaten“. Doch was genau ist darunter zu verstehen? Ziel war die Entwicklung eines Programmpakets in SAS, das die automatische Erstellung von Ergebnistabellen im Rahmen der Auswertung klinischer Studien ermöglicht. Im Rahmen einer Masterarbeit des Studiengangs Medical Biometry / Biostatics wurde hierfür in Kooperation mit dem pharmazeutischen Unternehmen Boehringer Ingelheim ein Konzept entwickelt, dessen Methodik und Implementierung im Folgenden näher beleuchtet werden.

Schlüsselwörter: Metadatengetriebene Programmierung, TLF, Metadaten

1 Einleitung

Eine wichtige Aufgabe im Rahmen der Statistischen Programmierung ist die Auswertung klinischer Studien. Dem Programmierer liegen hierfür Spezifikationen vor, anhand derer er die benötigten Analysedatensätze erstellt, die statischen Analysen durchführt und die Ergebnisse in Form von Tabellen, Listen und Grafiken (Tables, Listings, Figures, TFLs) für den klinischen Bericht aufbereitet. Die Spezifikationen werden dabei üblicherweise in Word- oder Excel-Dokumenten verwaltet und können schnell durch die Kopier- und Einfügen-Funktion in SAS-Programme eingefügt werden.

Dieser Ansatz bringt jedoch einige Nachteile mit sich (Dilorio, Abolafia, ohne Jahr): Je nach Umfang der zu erstellenden Tabellen kann es sehr mühselig sein, jede einzelne Spezifikation, sei es der Name einer Variable oder der Tabellentitel, in jedes SAS-Programm durch die Kopieren und Einfügen-Methode einzubinden. Darüber hinaus ist dieses Vorgehen sehr ineffizient. Sobald es zu einer einzigen Spezifikationsänderung kommt, muss diese in jedem SAS-Programm vorgenommen werden. Möglicherweise geht dabei der Überblick verloren und ein Programm wird nicht angepasst. Es nutzt stattdessen die ursprüngliche Spezifikation und die Ergebnisse werden infolgedessen nicht korrekt in der Tabelle dargestellt. Diese Problematik wird in Abbildung 1 veranschaulicht.

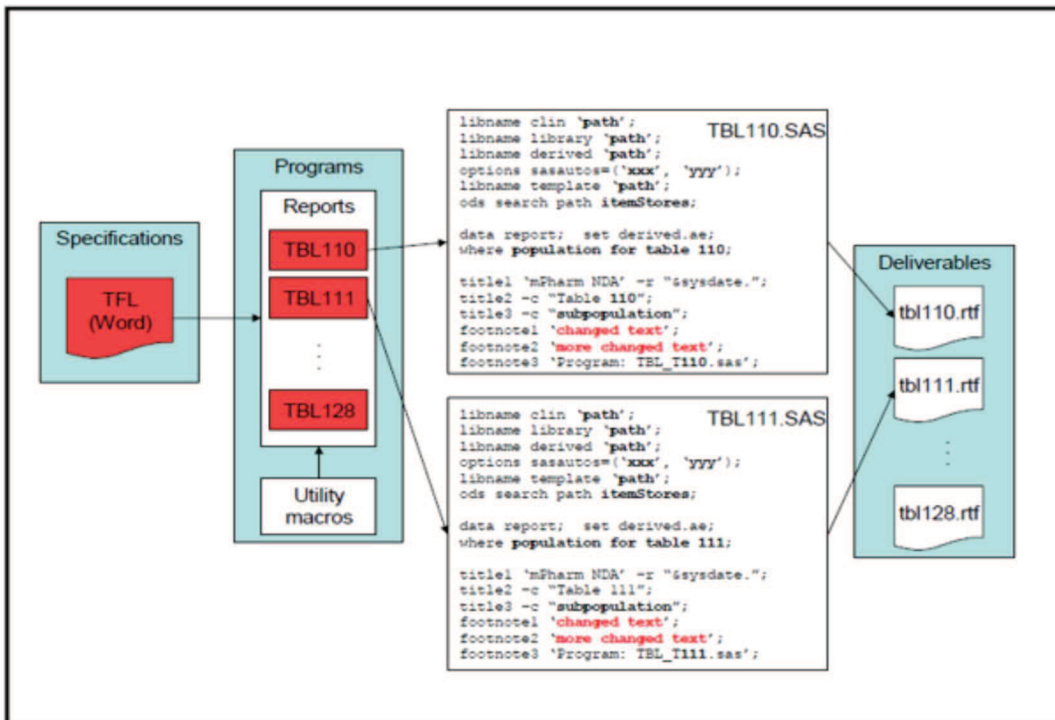


Abbildung 1: Änderung von Fußnoten (Quelle: Dilorio, Abolafia, ohne Jahr, S.3)

Dabei sollen für eine Studie zehn Tabellenprogramme programmiert werden. Die Tabellen sind hinsichtlich ihres Layouts identisch und unterscheiden sich nur im Titel und in Bezug auf die zugrundeliegende Population (siehe WHERE-Statement). Zehn Klone desselben Programms zu haben, wäre sehr ineffizient und würde zu den bereits geschilderten Problemen führen. Eine Änderung der Fußnoten, wie in Abbildung 1 dargestellt, würde beispielsweise eine Änderung aller Tabellenprogramme nach sich ziehen. Im schlimmsten Fall werden die Änderungen für einige Programme nicht übernommen und es kommt zu einer falschen Darstellung der Tabellen.

An diese Problematik setzt das für die Masterarbeit entwickelte Konzept an. Metadaten, die auf Spezifikationen basieren, sollen genutzt werden, um eine automatische Erzeugung von Ergebnistabellen für den Klinischen Bericht (CTR-Appendix-Tabellen) zu ermöglichen. Fester Programmcode soll durch dynamischen Programmcode ersetzt werden, um somit flexibler und robuster gegenüber Spezifikationsänderungen zu werden.

2 Anforderungen an das Konzept

Im Folgenden werden nun einige Anforderungen beschrieben, die bei der Implementierung des metadatengetriebenen Programmierkonzepts berücksichtigt werden sollten.

Einhaltung regulatorischer Anforderungen

Ein wichtiges Kriterium für das Konzept der Masterarbeit stellt die Einhaltung regulatorischer Anforderungen dar. Dazu zählt, dass sämtliche Analysen auf Basis CDISC ADaM-konformer Studiendaten erfolgen. Vor Beginn der Implementierung wurden da-

her die vorliegenden Daten in eine ADaM-konforme Datenstruktur transformiert. Darüber hinaus wurde für die Erfassung der Metadaten eine Umgebung geschaffen, die eine Verwaltung von Analysis Results Metadata berücksichtigt.

Erzeugung von vier verschiedenen CTR-Appendix-Tabellentypen

Mit Hilfe des metadatengetriebenen Ansatzes soll es möglich sein, alle CTR-Appendix-Tabellen für den Klinischen Bericht zu erzeugen. Für die Entwicklung und Implementierung des Konzeptes wurden die Tabellen hinsichtlich ihrer Darstellung und ihrer Analysemethoden analysiert und in verschiedene Tabellentypen eingeteilt, von denen vier Tabellentypen für die Implementierung selektiert wurden. Hierzu zählen:

- deskriptive Tabellen
- deskriptive Tabellen mit einer hierarchischen Ordnung
- Dispositionstabellen
- Efficacy-Tabellen, die das Mixed Model Repeated Measures (MMRM)-Modell nutzen

Benutzerfreundliche Eingabe und Verwaltung der Metadaten

Ein Kernstück des Konzeptes stellt die Eingabe und Verwaltung der Metadaten dar. Diese enthalten die Informationen, die benötigt werden, um eine CTR-Appendix-Tabelle zu erstellen. Aus diesem Grund sollte im Rahmen des Konzeptes eine Umgebung geschaffen werden, welche die benutzerfreundliche¹, manuelle Pflege sowie die elektronische Speicherung und Nutzung der Metadaten ermöglicht.

Möglichst geringer Programmieraufwand für den Anwender

Mit Hilfe eines metadatengetriebenen Ansatzes soll die Effizienz im Rahmen der Tabellenprogrammierung gesteigert werden. Aus diesem Grund wurde bei der Implementierung des Konzeptes eine Lösung angestrebt, die bei der finalen Anwendung möglichst wenig Programmieraufwand für den Benutzer erfordert, um eine CTR-Appendix-Tabelle zu erstellen.

3 Die Tabellentypen

Dieser Abschnitt beschreibt die vier Tabellentypen, die bei der Implementierung des metadatengetriebenen Konzeptes verwendet wurden. Dabei ist zu beachten, dass die Kategorisierung der CTR-Appendix-Tabellen auf Basis der Darstellung, Analysemethoden und Programmierung erfolgte und somit keine allgemeingültige Tabellen-Typisierung im Rahmen klinischer Studien darstellt. Die Darstellung der Ergebnisse der statistischen Analysen ist in den folgenden Beispieltabellen jeweils mit „x“ angegeben, wobei „x“ für jede natürliche Zahl von 0-9 steht.

¹ Benutzerfreundlich meint in diesem Zusammenhang die Verwaltung der Metadaten über eine übersichtliche Benutzeroberfläche sowie eine leichte Aktualisierbarkeit der Metadaten.

Deskriptive Tabellen (Typ 1)

Dieser Tabellentyp ermöglicht die Darstellung deskriptiver Statistiken für kategorielle und stetige Variablen. Kategorielle Variablen werden sowohl mit absoluten als auch mit relativen Häufigkeiten dargestellt. Für stetige Variablen können die statistischen Maßzahlen benutzerdefiniert angegeben werden. Dabei ist es möglich, dass unterschiedliche Maßzahlen für verschiedene Variablen verwendet werden. Beispielsweise kann der Benutzer festlegen, dass für Alter das arithmetische Mittel, der Median und die Standardabweichung dargestellt werden, während für das Gewicht neben diesen drei Maßzahlen auch die Varianz berücksichtigt wird. Darüber hinaus lässt sich für diesen Tabellentyp die Präzision, also die Anzahl an Nachkommastellen, der statistischen Maßzahlen steuern. Deskriptive Tabellen sind so aufgebaut, dass die Darstellung der jeweiligen Statistiken spaltenweise erfolgt. D.h. für jede Behandlungsgruppe werden die Statistiken in einer separaten Spalte dargestellt. Zusätzlich enthält der Tabellentyp eine weitere Spalte „Total“, in der die Ergebnisse ungruppiert zusammengefasst. Dieser Tabellentyp eignet sich insbesondere für die Darstellung der Auswertung demografischer Daten (vgl. Abbildung 2).

Table 190: Demographic data – treated set

	Placebo	New Drug 1	New Drug 2	Comparator	Total
Number of patients [N (%)]	xxx (100.0)	xxx (100.0)	xxx (100.0)	xxx (100.0)	xxx (100.0)
Gender [N (%)]					
Male	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Female	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Race [N (%)]					
American Indian / Alaska Native	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Asian	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Black / African American	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Native Hawaiian / Pacific Islander	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
White	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Ethnicity [N (%)]					
Not Hispanic / Latino	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Hispanic / Latino	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Age [years]					
N	xx	xx	xx	xx	xx
Mean	xx.x	xx.x	xx.x	xx.x	xx.x
SD	xx.x	xx.x	xx.x	xx.x	xx.x
Min	xx	xx	xx	xx	xx
Q1	xx	xx	xx	xx	xx
Median	xx.x	xx.x	xx.x	xx.x	xx.x
Q3	xx	xx	xx	xx	xx
Max	xx	xx	xx	xx	xx

Abbildung 2: Beispiel für eine deskriptive CTR-Appendix-Tabelle

Deskriptive Tabellen mit hierarchischer Ordnung (Typ 2)

Einer Sonderform für die Darstellung deskriptiver Statistiken nimmt dieser Tabellentyp ein. Im Gegensatz zum vorangegangenen Tabellentyp fasst diese Art von CTR-Appendix-Tabellen nur kategorielle Variablen zusammen. Die Darstellung erfolgt in einer hierarchischen Ordnung, die vom Benutzer festgelegt werden kann. Kategorielle Variablen werden sowohl mit absoluten als auch mit relativen Häufigkeiten dargestellt. Die Besonderheit besteht dabei darin, dass sich die relativen Häufigkeiten innerhalb einer Spalte nicht zwangsläufig zu 100% aufsummieren, wie dies beim vorangegangenen Tabellentyp der Fall ist. Ein Beispiel:

Innerhalb einer deskriptiven CTR-Appendix-Tabelle (Typ 1) zählt die Behandlungsgruppe „New Drug 1“ eine absolute Häufigkeit von 100 Patienten. Von diesen 100 Patienten sind 55 Patienten weiblich (55%), 45 männlich (45%). Addiert man beide relativen Häufigkeiten, so ergeben sich 100%.

Anders verhält es sich jedoch für deskriptive Tabellen mit hierarchischer Ordnung. Von den 100 Patienten der Behandlungsgruppe „New Drug 1“ wurde bei 5% der Patienten bereits Krebs diagnostiziert, 2% der Patienten leiden an Herzrhythmusstörungen, während 1% bereits einen Schlaganfall erlitten. Weitere Erkrankungen innerhalb dieser Gruppe wurden nicht diagnostiziert. Es wird offensichtlich, dass die 100% durch Addition der einzelnen relativen Häufigkeiten nicht erreicht werden. Die Nutzung dieses Tabellentyps ist insbesondere für die Auswertung der Krankengeschichte (Medical History) empfehlenswert (vgl. Abbildung 3).

Table 190: Medical history – treated set

	Placebo	New Drug 1	New Drug 2	Comparator	Total
Number of patients [N (%)]	xxx (100.0)	xxx (100.0)	xxx (100.0)	xxx (100.0)	xxx (100.0)
Number (%) of patients with history of					
Previous myocardial infarction	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Cerebrovascular accidents	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
...	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)

Abbildung 3: Beispiel für eine deskriptive CTR-Appendix-Tabelle mit hierarchischer Ordnung

Dispositionstabellen (Typ 3):

Dispositionstabellen zeigen eine Übersicht in Bezug auf den Einschluss eines Patienten in eine Studie („Enrolled“, „Not entered/randomized“, „Entered/randomized“, usw.) und das Absetzen der Studienmedikation. Auch dieser Tabellentyp besitzt eine hierarchische Ordnung. So wird zum Beispiel das vorzeitige Absetzen der Studienmedikation (Prematurely discontinued from trial medication) gezählt, wenn eines der folgenden Ereignisse eintritt:

- Unerwünschte Ereignisse (Adverse Events)
- Fehlende Wirksamkeit (Lack of efficacy)
- Verletzungen des Studienprotokolls (Non compliant with protocol)
- Nachverfolgung eines Patienten nicht mehr möglich (Lost to follow-up)
- Einwilligung aufgrund unerwünschter Ereignisse widerrufen (Consent withdrawn not due to adverse event)

Darüber hinaus sind Adverse Events in drei weitere Ereignisse unterteilt.

Die Aufsummierung einzelner relativer Häufigkeiten innerhalb einer Behandlungsgruppe ergibt immer 100%, wobei die Summe relativer Häufigkeiten von Ereignissen einer Ebene immer der relativen Häufigkeit der nächsthöheren Ebene entsprechen.

Ein Beispiel: Innerhalb einer Studie umfasst die Behandlungsgruppe „New Drug 1“ 100 Patienten. Von diesen 100 Patienten führte bei 20 Patienten ein Adverse Event dazu, die Behandlung vorzeitig abzubrechen. Dies ergibt eine relative Häufigkeit von 20%. Von den 20 Adverse Events sind 10 darauf zurückzuführen, dass sich die Krankheit im Verlauf der Studie weiter verschlimmert hat. Bei 5 der 20 Fälle hat sich eine bereits bestehende Krankheit weiter verschlechtert, während weitere 5 Fälle der Kategorie „Other adverse events“ zugeordnet werden. Die relativen Häufigkeiten ergeben sich aus den 100 Patienten der Behandlungsgruppe:

- Worsening of disease under study (10%)
- Worsening of other pre-existing disease (5%)
- Other adverse events (5%)

Die Summe dieser relativen Häufigkeiten ergibt die relative Häufigkeit der nächsthöheren Ebene (Adverse Events, 20%). Abbildung 4 zeigt ein Beispiel für eine Dispositionstabelle.

Table 200: Disposition of patients

	Placebo N (%)	New Drug 1 N (%)	New Drug 2 N (%)	Comparator N (%)	Total N (%)
Enrolled					xxx
Not entered/randomised					xxx
Entered/randomized	xxx	xxx	xxx	xxx	xxx
Not treated	xxx	xxx	xxx	xxx	xxx
Treated	xxx (100.0)	xxx (100.0)	xxx (100.0)	xxx (100.0)	xxx (100.0)
Not prematurely discontinued from trial medication	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Prematurely discontinued from trial medication	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Adverse event	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Worsening of disease under study	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Worsening of other pre-existing disease	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Other adverse event	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Lack of efficacy	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Non compliant with protocol	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Lost to follow-up	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Consent withdrawn not due to adverse events	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Other	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)

Abbildung 4: Beispiel für eine CTR-Appendix-Tabelle zur Darstellung der Patientendisposition

Efficacy-Tabellen (Typ 4)

Dieser Tabellentyp ermöglicht die Darstellung von Endpunktanalysen auf Basis gemischter Modelle mit Messwiederholungen (Mixed Model for Repeated Measures). Gemischte Modelle ermöglichen die Modellierung von individuen- bzw. clusterspezifischen Effekten. Sie finden insbesondere dann Anwendung, wenn die Anzahl der Beobachtungen der einzelnen Individuen sehr klein ist. Das Modell kann sowohl feste, kategorische Effekte (z.B. Behandlung, Zentrum, Interaktion zwischen Behandlung und Visite) als auch feste, stetige Kovariablen (z.B. Baseline, Interaktion zwischen Baseline und Behandlung) umfassen, die jeweils vom Benutzer angegeben werden. Die Patienten werden als zufällige Effekte (random effects) mit einbezogen. Die verwendete Kovarianzstruktur wird benutzerdefiniert festgelegt. Eine Beispieltabelle für diesen Tabellentypen ist in Abbildung 5 dargestellt.

Table 40: <Endpoint name> [unit] - MMRM results – FAS

Timepoint	Treatment	N	Adjusted* mean (SE)	----- Active vs Placebo -----		
				Adjusted* mean of difference (SE)	xx% CI	p-value
Week 4	Placebo	xx	xx.x (xx.xx)			
	New Drug 1	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	New Drug 2	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	Comparator	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
Week 8	Placebo	xx	xx.x (xx.xx)			
	New Drug 1	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	New Drug 2	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	Comparator	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
Week 16	Placebo	xx	xx.x (xx.xx)			
	New Drug 1	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	New Drug 2	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	Comparator	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
Week 24	Placebo	xx	xx.x (xx.xx)			
	New Drug 1	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	New Drug 2	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx
	Comparator	xx	xx.x (xx.xx)	xx.x (xx.xx)	(xx.x , xx.x)	x.xxxx

Abbildung 5: Beispiel für deskriptive CTR-Appendix-Tabelle zur Darstellung von Efficacy-Daten

4 Methodik

4.1 Identifizierung und Kategorisierung der Metadaten

Für die Entwicklung einer metadatengetriebenen Anwendung war zunächst eine Identifizierung der benötigten Metadaten erforderlich. Um ein besseres Verständnis über die benötigten Informationen zu erlangen, wurden die verschiedenen CTR-Appendix-Tabellen in Ihrer Form und Struktur detailliert analysiert, und, basierend auf der statistischen Analysemethode, der Strategie der Programmierung pro Tabelle und der Art der Darstellung der Ergebnisse in den Tabellen, in verschiedene Tabellentypen eingeteilt. Von diesen wurden vier Tabellentypen für diese Arbeit selektiert. Die Zerlegung der Tabellentypen in einzelne Komponenten ermöglichte dabei eine Einteilung der benötigten Informationen in die Kategorien „outputspezifisch“, „displayspezifisch“ und „analy-sespezifisch“.

Outputspezifisch []

Hierbei handelt es sich um Metadaten, die eine CTR-Appendix-Tabelle spezifizieren. Ein Großteil dieser Metadaten wird bereits im ToC dokumentiert (z.B. Titel, Fußnoten, Tabellenummer)

Displayspezifisch []

Diese Metadaten steuern das Aussehen einer CTR-Appendix-Tabelle. Hierzu zählen beispielsweise Schriftart und Schriftgröße, die Spaltenbreite oder Leerzeilen innerhalb einer CTR-Appendix-Tabelle.

Analysespezifisch []

Analysespezifische Metadaten enthalten Informationen in Bezug auf die statistische Analyse. Beispiele für diese Kategorie stellen Analysevariablen oder Maßzahlen stetiger Variablen dar.

Die nachfolgende Abbildung veranschaulicht die verschiedenen Kategorien für eine Beispieltabelle:

	<Placebo>	<Tio R2.5>	<Tio R5>	<Salmeterol>	<Total>
Table 190: Demographic data - treated set					
Number of patients [N (%)]	xxx (100.0)	xxx (100.0)	xxx (100.0)	xxx (100.0)	xxx (100.0)
Gender [N (%)]					
Male	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Female	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Race [N (%)]					
American Indian / Alaska Native	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Asian	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Black / African American	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Native Hawaiian / Pacific Islander	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
White	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Ethnicity [N (%)]					
Not Hispanic / Latino	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Hispanic / Latino	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Age [years]					
N	xx	xx	xx	xx	xx
Mean	xx.x	xx.x	xx.x	xx.x	xx.x
SD	xx.x	xx.x	xx.x	xx.x	xx.x
Min	xx	xx	xx	xx	xx
Q1	xx	xx	xx	xx	xx
Median	xx.x	xx.x	xx.x	xx.x	xx.x
Q3	xx	xx	xx	xx	xx
Max	xx	xx	xx	xx	xx

Abbildung 6: Identifizierung von Metadaten

Nach diesem Schema wurden alle vier Tabellentypen analysiert. Die meisten Informationen, die zur Darstellung der statistischen Ergebnisse bei der Tabellenprogrammierung berücksichtigt werden müssen, sind in der Abbildung bereits grafisch zu erkennen. Doch es gibt darüber hinaus noch weitere Informationen, die für die inhaltlich korrekte statistische Analyse und Darstellung der Ergebnisse eine wichtige Rolle einnehmen. Um

diese Elemente zu identifizieren, wurden anschließend an die Analyse der vier vorgestellten Tabellentypen ausgewählte CTR-Appendix-Tabellen gemäß CTR display plan mit SAS programmiert. Die entwickelten SAS-Programme wurden nun detailliert analysiert, um weitere Metadaten zu identifizieren. Anhand der Programmcodes konnten schließlich weitere Bestandteile bzw. Elemente, die sich im SAS-Programm häufig wiederholen sowie durch festen Programmcode in Ihrer Dynamik und Flexibilität eingeschränkt sind, identifiziert werden. Diese Bestandteile wurden gemäß der vorgenommenen Metadatenkategorisierung in outputspezifische, displayspezifische und analysespezifische Metadaten eingeteilt werden.

4.2 Erstellung einer Umgebung zur elektronischen Speicherung und Pflege der Metadaten

Die Basis für eine automatische Erzeugung einer CTR-Appendix-Tabelle liegt in der Eingabe und Verwaltung der identifizierten Metadaten. Zum einen wurde hierfür der MetadataToC zur elektronischen Speicherung und Pflege der Metadaten entwickelt (vgl. Kapitel 5.1), zum anderen der bestehende Boehringer Ingelheim (BI)-interne ADS-Plan geringfügig erweitert (vgl. Kapitel 5.2).

Die Kombination beider Dokumente ermöglicht die vollständige Erfassung aller benötigten Metadaten für eine automatische CTR-Appendix-Tabellen-Erstellung.

4.3 Programmatische Implementierung

Die Implementierung des Konzeptes erfolgte mit dem statistischen Softwarepaket SAS und wurde durch die Entwicklung des Dynamic Tools (vgl. Kapitel 6.1) sowie durch die Programmierung von tabellentypspezifischen Makros (vgl. Kapitel 6.2) realisiert. Das Konzept lässt sich dabei wie folgt beschreiben:

Jede CTR-Appendix-Tabelle wird durch ein tabellentypspezifisches Makro erzeugt, d.h. für jeden Tabellentyp wurde ein separates Makro entwickelt, das die statistische Analyse und die anschließende Erzeugung einer CTR-Appendix-Tabelle ermöglicht. Die einzelnen tabellentypspezifischen Makros benötigen, in Abhängigkeit des Tabellentyps, unterschiedliche Parameter für den Makroaufruf. Für die Erzeugung der gewünschten CTR-Appendix-Tabellen ist es nun erforderlich, eine automatisierbare Verknüpfung zwischen den verwalteten Metadaten und jenen Makroparametern herzustellen.

Hierfür wird das Dynamic Tool genutzt, das neben dem MetadataToC aufgrund seiner umfangreichen Funktionalitäten das zweite Kernstück des Konzeptes darstellt. Diese umfassen:

- den Import der Metadaten in SAS,
- die Erzeugung permanenter SAS-Datensätze mit den definierten Metadaten und SAS-Formaten,
- verschiedene Arten von Fehler-Checks sowie

- die Generierung von Analyseprogrammen, die den benötigten Makroaufruf für die Erstellung einer CTR-Appendix-Tabelle beinhalten und die hierfür erforderlichen Metadaten dem Makro als Makroparameter-Werte übergibt.

Anhand der importierten Metadaten stellt das Dynamic Tool eine Verknüpfung zum tabellentypspezifischen Makro her. Das heißt es prüft für jede CTR-Appendix-Tabelle, welches Makro für die Erstellung notwendig ist und schreibt den dafür benötigten Makroaufruf in ein Analyseprogramm. Mit dem Ausführen des Analyseprogramms wird anschließend die CTR-Appendix-Tabelle erzeugt. Eine detaillierte Prozessbeschreibung folgt in Kapitel 7.

5 Dokumente zur Eingabe und Verwaltung der Metadaten

5.1 MetadataToC

Beim MetadataToC² handelt es sich um ein Excel-Dokument, das für die Eingabe sowie für die Verwaltung der identifizierten Metadaten entwickelt wurde. Es setzt sich aus dem MetadataToC-Sheet und dem CT-Sheet zusammen, die im Folgenden nun vorgestellt werden.

MetadataToC-Sheet

Das MetadataToC-Sheet fasst alle Informationen zusammen, die in Bezug auf die Erstellung einer CTR-Appendix-Tabelle benötigt werden. Dabei erfolgt die Verwaltung der Metadaten spaltenweise. Des Weiteren orientiert sich Aufbau des MetadataToC-Sheets sich an einer zeilenweisen Struktur, d.h. pro zu erzeugende CTR-Appendix-Tabelle wird eine Zeile im MetadataToC-Sheet angelegt. Tabelle 1 zeigt den strukturellen Aufbau des MetadataToC-Sheets anhand ausgewählter Metadaten für zwei CTR-Appendix-Tabellen.

Tabelle 1: Ausschnitt vom MetadataToC-Sheet des MetadataToCs

STUDY	APPENDIX	NUMBER	...	TABTYPE	DATATYP
1234 5678	15.1	1: 1	...	disp	ADSL
1234 5678	15.1	4: 1	...	demo	ADSL

StatCat-Sheet

Mit Hilfe des StatCat-Sheets kann für deskriptive CTR-Appendix-Tabellen (Typ 1) festgelegt werden, welche Maßzahlen bei der Darstellung stetiger Variablen verwendet werden. Hierfür wird eine benutzerdefinierte Kategorie (StatCat) angelegt, die für mehrere Variable verwendet werden kann.

² ToC = Table of Contents

Neben den darzustellenden Maßzahlen lässt sich auch die Präzision³ und die Reihenfolge der Maßzahlen innerhalb einer CTR-Appendix-Tabelle spezifizieren. Ein Beispiel für die Verwendung des StatCat-Sheets ist in Tabelle 2 gegeben.

Tabelle 2: Ausschnitt vom StatCat-Sheet des MetadataToCs

STATCAT	STATISTICS	FORMAT	ORDER
FA_1	n	4.0	1
FA_1	mean	5.1	2
FA_1	std	5.1	3
FA_1	min	4.0	4

5.2 ADS-Plan

Die eigentliche Idee eines ADS-Plans besteht darin, zu erzeugende Analysedatensätze sowie alle darin enthaltene Analysevariablen durch Metadatenpezifikationen exakt zu definieren und zu beschreiben. Im Rahmen dieses metadatengetriebenen Programmkonzepts wird der ADS-Plan darüber hinaus genutzt, um Variablen innerhalb einer CTR-Appendix-Tabelle zu beschriften sowie Controlled Terms für eine Variablendekodierung zu erstellen.

6 Programmatische Nutzung der Metadaten

6.1 Dynamic Tool

Beim Dynamic Tool handelt es sich um das entwickelte SAS-Makro %dynamic zur programmatischen Nutzung der eingegeben bzw. verwalteten Metadaten. Mit Hilfe des Dynamic Tools werden die Metadaten in SAS importiert und in Form permanenter SAS-Datensätze gespeichert, die für die spätere statistische Analyse genutzt werden. Anhand der importierten Metadaten prüft das Dynamic Tool, welches tabellentypspezifische Makro⁴ für die Erstellung der CTR-Appendix-Tabelle benötigt wird und schreibt die erforderliche Syntax für den Makroaufruf des identifizierten Makros in ein neues SAS-Programm (Analyseprogramm, vgl. Abschnitt „Generierung des Analyseprogramms“). Die Funktionalitäten des %dynamic-Makros lassen sich grob in die folgenden sechs Punkte gliedern:

³ Präzision = Anzahl der Nachkommastellen

⁴ Pro Tabellentyp wurde ein separates Makro entwickelt, das die statistische Analyse sowie die Erstellung der entsprechenden CTR-Appendix-Tabelle ermöglicht. Eine Definition der jeweiligen Makros folgt im Kapitel 4.6.1

Tabelle 3: Beschreibung der Funktionalitäten des Dynamic Tools

Funktionalität	Beschreibung
„Aufräumen“ der SAS-Umgebung	Ermöglicht die korrekte Ausführung des Dynamic Tools. Löschen aller Datensätze der Work-Library sowie existierender Makrovariablen.
Import der Metadaten	Import der eingegebenen Metadaten im MetadataToC und ADS-Plan
Erstellung von SAS-Formaten	Erzeugung von SAS-Formaten auf Basis der im ADS-Plan spezifizierten Controlled Terms
Generierung des Analyseprogramms	Erzeugung eines Analyseprogramms auf Basis der eingegebenen Metadaten im MetadataToC und ADS-Plan. Erstellung der CTR-Appendix-Tabelle erfolgt durch Ausführen des Analyseprogramms
Fehlerchecks	Implementiere Fehlerchecks prüfen unplausible Eingaben im MetadataToC

Generierung des Analyseprogramms

Die Generierung des Analyseprogramms basiert auf der Verwendung des PUT-Statements, mit Hilfe dessen vorgegebener Programmcode in das Analyseprogramm geschrieben wird. Erforderlich für die Generierung des Analyseprogramms ist der Aufruf der drei Makros %head1, %parameters und %body (vgl. Abbildung 7). Dabei wird der in Abbildung 7 dargestellte Programmcode für jeden vom Dynamic Tool erzeugten OUTTABLE-Datensatz⁵ durchlaufen. Dies ermöglicht die Generierung eines Analyseprogramms pro Tabellentyp.

```

*** WRITE TABLE-SPECIFIC SYNTAX INTO A NEW PROGRAM;
data tempcall;
  file "C:\masterarbeit\1st\ctr\&_study.\&_program";
  set metadata.outtable_&_type. end = eof;
  %head1;  *** WRITE PROGRAM HEADER;
  put " ";
  put "/******";
  put "** " Appendix " " Number " " Type;
  put "** " title_1;
  put "** " title_2;
  put "*****/";
  put " ";

  %parameters;  *** WRITE TABLE-SPECIFIC SYNTAX FOR MACRO CALL;
  put " ";
  %body;      *** WRITE "END-OF-PROGRAM-FOOTER";
run;

```

Abbildung 7: Generierung des Analyseprogramms

⁵ Pro Tabellentyp wird vom Dynamic Tool ein permanenter tabellentypspezifischer Datensatz (OUTTABLE) erzeugt, der die entsprechenden Metadaten verwaltet.

- `%head1`:
Erzeugt den Programmkopf bzw. Header des Analyseprogramms. Der Header beinhaltet dabei Angaben zur Erstellung des Analyseprogramms sowie spezifische Informationen über die zu erstellende CTR-Appendix-Tabelle.
- `%parameters`:
Erzeugt den Programmcode für den tabellentypspezifischen Makroaufruf, der für die Erstellung einer CTR-Appendix-Tabelle benötigt wird. Die Identifizierung des benötigten tabellentypspezifischen Makroaufrufs wurde programmatisch dabei wie folgt umgesetzt:
Innerhalb des Makros `%parameters` wurde für jedes tabellentypspezifische Makro ein separater Makroaufruf definiert. Ein Beispiel hierfür ist in Abbildung 8 dargestellt. Wird das Dynamic Tool später vom Benutzer ausgeführt, so erfolgt über `TABTYPE`⁶ die Identifizierung des benötigten Makroaufrufs und die damit verknüpfte Syntax wird in das Analyseprogramm geschrieben.

```

%macro parameters;
  if upcase(tabtype) = "DEMO" then do;
    put " ";
    put "%demo(";
    put "      ds          = " dataset;
    put "      , number   = " number;
    put "      , appendix  = " appendix;
    put "      , popu      = " popuset;
    put "      , varlist   = (" analvar)";
    put "      , treatment = (" treatord)";
    if not missing(byvar) then put "      , byvar      = " byvar;
    if not missing(where) then put "      , where     = " where;
    /* optional footnotes [write in macro call, if used];
    %do z = 6 %to 10;
      if not missing(foot&z.) then put "      , foot&z. = " foot&z.;
    %end;
    put "      );";
  end;

```

Abbildung 8: Makroaufruf, der für das Makro `%demo` erzeugt wird

- `%body`:
Erzeugt einen Kommentarblock mit dem Hinweis „END OF PROGRAM“ und kennzeichnet damit das Ende des Analyseprogramms.

6.2 Tabellentypspezifische Makros

Die Erstellung der CTR-Appendix-Tabellen basiert auf den tabellentypspezifischen Makros, die im Rahmen des Konzepts implementiert wurden. Hierfür wurde pro Tabellentyp (vgl. Kapitel 3) ein separates Makro entwickelt, das die statistische Analyse so-

⁶ `TABTYPE` = Tabellentyp der CTR-Appendix-Tabelle

wie die Erstellung der entsprechenden CTR-Appendix-Tabelle ermöglicht. Hierzu zählen die folgenden vier Makros:

%demo:

Das Makro %demo ermöglicht die Erzeugung von CTR-Appendix-Tabellen zur Darstellung deskriptiver Analyseergebnisse stetiger und kategorieller Variablen. Gemäß der vorgenommenen Tabellentyp-Kategorisierung sind die zu erzeugenden CTR-Appendix-Tabellen dem Tabellentyp 1 zugeordnet (vgl. Kapitel 3).

%medh:

Das Makro %medh ermöglicht die Erzeugung von CTR-Appendix-Tabellen zur Darstellung deskriptiver Analyseergebnisse kategorieller Variablen in einer hierarchischen Ordnung. Gemäß der vorgenommenen Tabellentyp-Kategorisierung sind die zu erzeugenden CTR-Appendix-Tabellen dem Tabellentyp 2 zugeordnet (vgl. Kapitel 3).

%disp:

Das Makro %disp ermöglicht die Erzeugung von CTR-Appendix-Tabellen zur Darstellung der Patientendisposition. Gemäß der vorgenommenen Tabellentyp-Kategorisierung sind die zu erzeugenden CTR-Appendix-Tabellen dem Tabellentyp 3 zugeordnet (vgl. Kapitel 3).

%mmrm:

Das Makro %mmrm ermöglicht die Erzeugung von CTR-Appendix-Tabellen zur Darstellung statistischer Ergebnisse von Endpunktanalysen auf Basis gemischter Modelle mit Messwiederholungen (Mixed Model for Repeated Measures). Gemäß der vorgenommenen Tabellentyp-Kategorisierung sind die zu erzeugenden CTR-Appendix-Tabellen dem Tabellentyp 4 zugeordnet (vgl. Kapitel 3).

Jedes der vier tabellentypspezifischen Makros benötigt für seinen Makroaufruf die Angabe bestimmter Makroparameter. Nach Ausführen des Dynamic Tools wird auf Basis der importierten Metadaten ein Analyseprogramm zur Erstellung der CTR-Appendix-Tabelle generiert. Dieses beinhaltet den jeweiligen Makroaufruf, der für die Ausführung des tabellentypspezifischen Makros erforderlich ist. Die Makroaufrufe der vier tabellentypspezifischen Makros können jedoch auch manuell, d.h. ohne Ausführung des Dynamic Tools, erstellt werden. Voraussetzung dafür ist das Vorhandensein der entsprechenden Datensätze sowie die syntaktisch korrekte Verwendung der Makroparameter. Ein beispielhafter Makroaufruf ist in Abbildung 9 dargestellt.

```
%demo(  
  ds          = ADSL  
  , number    = 4: 1  
  , appendix  = 15.1  
  , popu      = TRTFN  
  , varlist   = (#SEXN, #RACEN, AGE(FA_1), #AGEGR1N)  
  , treatment = (Placebo, New Drug 1, New Drug 2, Comparator 2)  
);
```

Abbildung 9: Makroaufruf des Makros %demo

7 Prozessbeschreibung

Im Folgenden wird nun der Prozess des hier vorgestellten metadatengetriebenen Programmansatzes beschrieben, um die Funktionalitäten und das Zusammenspiel einzelner Bestandteile des Konzepts zu verdeutlichen.

Zunächst erfolgt die Eingabe der Metadaten im MetataToC und im ADS-Plan. Beide Dokumente werden anschließend mit Hilfe des Dynamic Tools in SAS importiert, welches die eingelesenen Informationen in Form von permanenten SAS-Datensätzen speichert und SAS-Formate bildet. Im Anschluss an die Datensatz- und Formaterstellung führt das Dynamic Tool drei verschiedene Fehlerchecks in Bezug auf die eingegebenen Metadaten im MetadataToC durch. Werden während dieser Prüfung vom Dynamic Tool Fehler in der Eingabe festgestellt, so werden diese dem Anwender im Log-Fenster durch eine entsprechende Fehlermeldung angezeigt. Je nach Art des Fehlers müssen die Eingaben im MetadataToC ggf. korrigiert und der Einleseprozess erneut durchgeführt werden.

Auf Basis der importierten Metadaten generiert das Dynamic Tool im weiteren Verlauf ein Analyseprogramm zur Erstellung der im MetadataToC definierten CTR-Appendix-Tabelle. Dieses Analyseprogramm beinhaltet den jeweiligen Makroaufruf, der für die Ausführung des tabellentypspezifischen Makros erforderlich ist. Dabei werden die anzugebenden Makroparameter automatisch vom Dynamic Tool vervollständigt - weitere Programmiermaßnahmen seitens des Benutzers sind nicht erforderlich.

Mit der Ausführung des Analyseprogramms wird zunächst die statistische Analyse durchgeführt, indem das tabellentypspezifische Makro durchlaufen wird. Dieses benötigt für die Durchführung der Analyse zum einen den entsprechenden Analysedatensatz, zum anderen die vom Dynamic Tool erstellten Metadaten-Datensätze sowie die gebildeten SAS-Formate. Abschließend wird die CTR-Appendix-Tabelle als LST-Datei im vom Benutzer definierten Verzeichnis gespeichert. Eine grafische Prozessbeschreibung ist in Abbildung 10 gegeben.

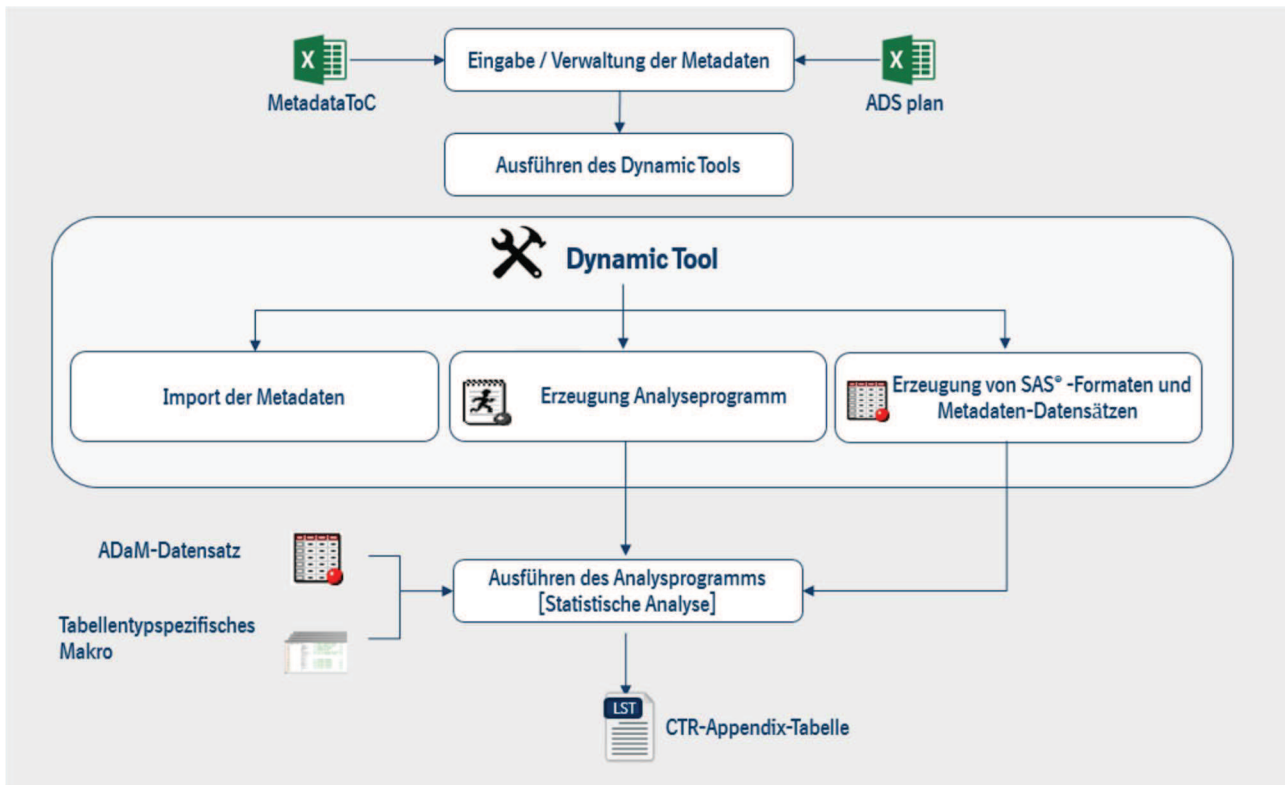


Abbildung 10: Prozess des metadatengetriebenen Programmierkonzepts

8 Validierung der Ergebnisse

Ein wichtiges Kriterium für eine Anwendung in der Praxis des vorgestellten metadatengetriebenen Programmierkonzepts ist die Gewährleistung, dass alle erzeugten CTR-Appendix-Tabellen die korrekten Ergebnisse der statistischen Analyse darstellen. Für die Sicherstellung valider Studienergebnisse durch das vorgestellte Programmpaket wurde die Strategie der Validierung durch Doppelprogrammierung gewählt. Dazu wurden vier CTR-Appendix-Tabellen einer Asthma-Studie sowohl ohne als auch mit Nutzung des metadatengetriebenen Programmpakets erzeugt.

Beide Ansätze lieferten für die vier erzeugten CTR-Appendix-Tabellen die gleichen Ergebnisse und waren identisch zu den originalen Studienergebnissen.

Der Beitrag basiert auf der Masterarbeit des Autors. Bei Interesse kann die vollständige Arbeit vom Autor angefordert werden.

Literatur

- [1] Dilorio, F., Abolafia, R. (ohne Jahr). The Design and Use of Metadata: Part Fine Art, Part Black Art. SUGI 31. Paper 104