

# Auswertung klinischer Studien mit SAS/PH-Clinical

Matthias Wahl

Schwarz Biosciences GmbH

Alfred Nobel Str. 10

40789 Monheim

matthias.wahl@schwarzpharma.com

## Zusammenfassung

SAS/PH-Clinical ist ein AF-basiertes add-on Modul für SAS, das es -einfach umschrieben- ermöglicht, Macros mit einem Graphical User Interface (GUI) auszustatten. Neben der erhöhten Anwenderfreundlichkeit des GUI sind PH-Clinical Templates aber auch wesentlich flexibler als herkömmliche Macros. Dieser Artikel gibt eine Übersicht über die Funktionen von PH-Clinical, erläutert, warum es zur Auswertung klinischer Studien besser geeignet ist als nur Base SAS und stellt das Gesamtkonzept von Schwarz Biosciences vor, in dem PH-Clinical zur Analyse klinischer Studien standardmäßig verwendet werden wird.

## 1 Einführung in SAS/PH-Clinical

Die einleitende Beschreibung, PH-Clinical sei im Prinzip „Macros mit GUI“ ist sehr simplifiziert. In Wirklichkeit ist PH-Clinical wesentlich komplexer und vereint drei wichtige Funktionsblöcke in einem Programm. Vorweg sei erwähnt, dass PH-Clinical eine Benutzerverwaltung hat, in der individuelle Benutzerrechte und Benutzerprofile vergeben werden können. Außerdem werden alle Zugriffe in einem Audit Trail protokolliert. Es ist sogar möglich, Änderungen auch nach „Überspeichern“ rückgängig zu machen. Die Ausstattung von PH-Clinical ist mit der eines Datenmanagementsystems vergleichbar. Für die Verwendung in der pharmazeutischen Industrie ist es wichtig, dass PH-Clinical die sog. CFR-Part-11 Anforderungen der amerikanischen Gesundheitsbehörde FDA erfüllt und somit im Falle eines behördlichen Audits keinerlei Probleme verursachen würde.

## 1.1 Studienverwaltung

Der erste Funktionsblock von PH-Clinical ist eine Datenbank-ähnliche Verwaltung von einzelnen SAS Datasets als sog. Studienobjekte. Mehrere SAS Datensätze werden zu einer Studie zusammengefasst und mit Informationen über die Datenstruktur mit einem getrennten Modul (Autoload Manager) nach PH-Clinical importiert. Man kann sich diesen Vorgang so vorstellen, dass gedanklich alle Datensätze der Studie zu einem riesigen Datensatz gemerged werden. Neben den Informationen, wo die Datensätze liegen und welche Datensätze zu einem Studienobjekt zusammengefasst werden sollen, muss PH-Clinical daher auch noch mitgeteilt werden, welche Variablen Merge-keys sind und welcher Hierarchie diese keys unterliegen. Außerdem können keinerlei strukturelle Änderungen mehr vorgenommen werden, wenn ein Studienobjekt einmal nach PH-Clinical importiert wurde, denn PH-Clinical speichert alle Informationen über die Variablen eines Studienobjektes (Name, Label, Format, etc). Dies ist insbesondere deshalb relevant, weil zum Zeitpunkt des Imports alle Variablen, die zur Auswertung benötigt werden, verfügbar sein müssen. Es ist zwar theoretisch denkbar, dass solche Variablen zum Analysezeitpunkt per Programm/Template erzeugt werden, aber in diesem Fall würde es sich um temporäre Variablen handeln, die nur zur Laufzeit dieses einen Templates verfügbar wären. Insofern stellt PH-Clinical hohe Ansprüche an die Datenqualität, sowohl in inhaltlicher, als auch in struktureller Sicht. Im Gegenzug wird die Analyse erleichtert, da unabhängig von der konkreten Studie eine normierte Datenform angenommen werden kann. Falls studienübergreifende Variablen immer die gleichen Attribute haben (Name, Label, Format, etc.), ist eine Projekt-bezogene Analyse mehrerer Studien in einer gepoolten Analyse kein Problem mehr. Es sei noch erwähnt, dass es sich hierbei nur um eine Datenbank-ähnliche Funktionalität handelt. PH-Clinical verwaltet nur strukturelle Informationen über die Studienobjekte, nicht *inhaltliche*. Die Grundlage zu einem Studienobjekt bilden immer noch einfache SAS Datensätze, über die lediglich Meta-Daten gespeichert werden und, solange diese Meta-Daten unverändert bleiben, behält das Studienobjekt unter PH-Clinical seine Gültigkeit. Es ist also theoretisch möglich, per Base SAS inhaltliche Änderungen in den SAS Datensätzen vorzunehmen, ohne das PH-Clinical dies „bemerkt“. Obwohl PH-Clinical nicht den Anspruch erhebt, ein Datenmanagementsystem zu sein, ist dies ein Schwachpunkt des Systems, der eine erhöhte Überwachung der Zugriffsrechte auf Betriebssystem-Ebene erfordert, aber erstens ist dies unter Base SAS nicht anders und zweitens kann dies, wie im folgenden gezeigt wird, auch vorteilhaft genutzt werden.

## 1.2 Studien Review

Der zweite Funktionsblock ist das sog. Browsing-Tool, welches es ermöglicht, Studiendaten anzusehen und einfache Statistiken (Diagramme, Häufigkeiten, Summary Statistics) per Mausklick zu erzeugen. Bei der Auswertung klinischer Studien ist dieses Tool nur limitiert zur explorativen Datenanalyse hilfreich, aber da, wie eben erwähnt, Änderungen in den Daten von PH-Clinical nicht wahrgenommen werden, solange die Struktur des Studienobjekts gewahrt bleibt, kann das Browsing Tool in der Praxis als Review-Tool für die klinischen Studien verwendet werden:

Sobald die ersten Patientendaten vorliegen (und hier reicht ein einziger vollständiger Patient), werden die Datensätze zu einem Studienobjekt gebündelt und nach PH-Clinical importiert. Mitarbeiter aus Monitoring, Trial Management und Drug Safety erhalten Leserechte auf die Studie und Anwendungsrechte für das Browsing-Tool. Es ist nun für diese Mitarbeiter möglich, die Studiendaten einzusehen und auf diese Art und Weise Fehler und Ungereimtheiten zu entdecken. Dadurch, dass mehr Daten (also zusätzliche Zeilen in den Ursprungs-Datensätzen) keine strukturelle Veränderung darstellen, kann das Studienobjekt sehr einfach aktualisiert werden, indem einfach die zugrunde liegenden Datensätze aktualisiert werden. Mit ein paar einfachen Templates, die aufgabenspezifisch für diese Anwender programmiert und freigeschaltet werden (z.B. Listing der Drug-related Adverse Events für Mitarbeiter der Drug Safety) wird PH-Clinical so zu einem Review-Tool, welches auch von SAS-unerfahrenen Anwendern nach kurzer Schulung benutzt werden kann. Dadurch, dass so frühzeitig ein großer Teil der involvierten Mitarbeiter in die Daten Einsicht erhält, erhoffen wir uns schnelleres Erkennen und damit Verbessern von hauptsächlich inhaltlichen Fehlern und somit langfristig eine bessere Datenqualität sowie ein besseres Verständnis für die Daten selbst.

## 1.3 Studienanalyse

Die Hauptfunktionalität (aus Sicht der statistischen Analyse von klinischen Studien) von PH-Clinical liegt aber im Programmiereteil. Es stehen alle unter Base SAS lizenzierten Module zur Verfügung. Die Programme/Templates unter PH-Clinical unterscheiden sich allerdings von Programmen unter Base SAS. In PH-Clinical sind die Templates in der Regel nicht mehr so spezifisch wie in Base SAS, sondern, da die Studienobjekte eine normierte Struktur haben, abstrakter und damit universeller einsetzbar. Am besten zu vergleichen sind die Templates mit Macros. Ein Template ist im Prinzip eine Mischung aus Base SAS Programm und Macro (alle Base SAS Programme würden auch unter PH-Clinical als Template laufen). Allerdings kann der Programmierer unabhängig von konkreten Daten programmieren. Hierzu werden statt den konkreten Datensätzen und Variablen nur Platzhalter – sog. Substitution fields –

verwendet, die erst zur Laufzeit durch entsprechende Auswahl konkretisiert werden. Genau wie Macro Variablen werden diese Platzhalter im Programm mit einem & gekennzeichnet. Es stehen verschiedene Typen von Platzhaltern zur Auswahl (einzelne Variablen, numerische Werte, Textfelder, spezielle GUI-spezifische Objekte, wie Tick Boxen oder klassische Macro Variablen). Für jedes im Programmtext mit & gekennzeichnete Feld muss anschließend die Bedeutung eingegeben werden. Die meisten (außer Studiendaten und klassischen Macro Variablen) dieser Objekte haben auch ein zugehöriges GUI-Objekt, welches in einem oder mehreren Benutzerfenstern angelegt werden muss, damit es zur Laufzeit vom Anwender z.B. mit einer Variable per Mausklick belegt werden kann. Ein Template kann programmtechnisch in mehrere Unterabschnitte unterteilt werden, um so z.B. bestimmte Aktionen von weiteren Eingaben des Anwenders abhängig zu machen. Neben User access control und Version control ist die Unabhängigkeit von speziellen Studien und damit die Flexibilität der Templates ein enorm grosser Vorteil. Macros unter Base SAS haben zwar eine ähnliche Flexibilität, sind aber für den Anwender wesentlich unfreundlicher zu bedienen. Am Beispiel eines Templates zur Erstellung von Summary Statistics (N, Mean, Standard Abweichung, Minimum, Median, Maximum) wird gezeigt, wie Templates programmiert werden und wie flexibel der Einsatz ist.

In der Analyse von klinischen Studien werden alle stetigen Variablen mit den oben erwähnten Summary Statistics zusammengefasst. Innerhalb einer Studie fallen solche Variablen in den verschiedensten Bereichen und damit in der Regel auch in verschiedenen Datensätzen an. Selbst wenn der Kern eines solchen Programms (oder falls eine Macro library zur Verfügung steht, das summary statistics Macro) für jeden Datensatz bzw. jede Variable übernommen werden kann, so sind doch am Ende viele verschiedene Base SAS Programme nötig, um die entsprechenden Tabellen einer Studiauswertung zu erstellen. In PH Clinical wird hierfür lediglich ein (wenn auch anspruchsvolles) Template benötigt.

Unser Schwarz Biosciences Summary Statistics Template besteht aus 3 Eingabefenstern und 3 Code Sections. Die verwendeten Procedures sind im wesentlichen PROC UNIVARIATE zur Berechnung der Summary Statistics, PROC REPORT für die formatierte Ausgabe der Ergebnisse, sowie notwendige Macros. Das Template läuft wie folgt ab:

- Wenn ein Studienobjekt geöffnet ist, wird dieses für die Analyse herangezogen, ansonsten wird der Nutzer aufgefordert, ein Studienobjekt festzulegen (wird von PH-Clinical automatisch erledigt).
- In einem ersten Fenster wird der User aufgefordert, die Analysevariablen (beliebig viele) aus der geöffneten Studie auszuwählen, sowie festzulegen, ob eventuell nach By-Variablen gruppiert werden soll (ebenfalls beliebig viele möglich). Außerdem sind noch einige nicht essentiell wichtige Festlegungen in Bezug auf Missing Values und Spaltenbreite bei der Ausgabe einstellbar, die aber in der Regel mit den voreingestellten Werten gut abgedeckt sind.

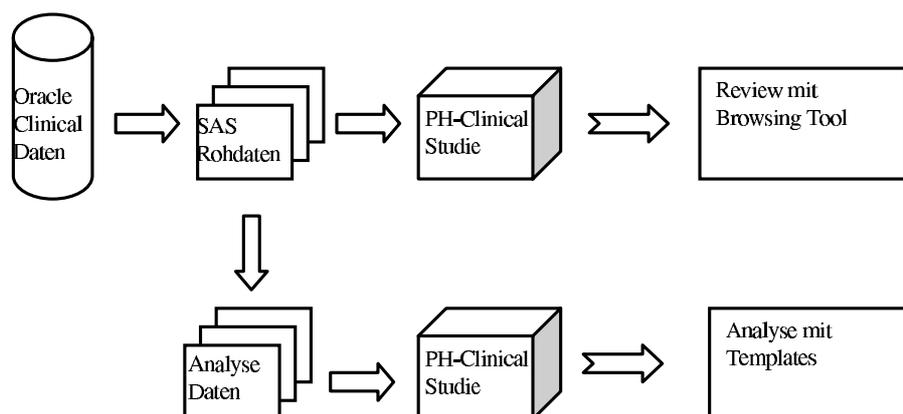
- Wenn der User seine Auswahl bestätigt, läuft eine Code Section, in der die ausgewählten By-Variablen ausgelesen werden und für das folgende Fenster vorbereitet werden.
- In einem zweiten Fenster, das automatisch geöffnet wird, wird der User aufgefordert zu bestimmen, ob und für welche der ausgewählten By-Variablen Totals berechnet werden sollen
- Nach Bestätigung läuft wieder eine Code Section, in der mit PROC UNIVARIATE die Summary Statistics berechnet werden.
- In einem letzten Fenster, das sich auch automatisch öffnet, kann der User nun das genaue Layout der Tabelle bestimmen. Dabei steht es ihm frei, welche Information in Spalten und welche in Zeilenrichtung angezeigt werden soll.
- Die letzte Code Section produziert nun mittels PROC REPORT die gewünschte Tabelle und zeigt diese, wie gewohnt, in einem Ausgabefenster an.

Dieses Template kann beliebig oft mit verschiedensten Auswahlmöglichkeiten laufen, wobei jeder Durchlauf, wie gewohnt, in einem Logfile protokolliert wird. In diesem Logfile sind die Platzhalter dann durch ihre aktuellen Belegungen ersetzt, so dass der Log wie der eines normalen Base SAS Programms aussieht. Mit einem Template können also fortan alle Summary Statistics Tabellen mit sehr wenigen und einfachen interaktiven Eingaben erstellt werden. Die Möglichkeit, mit einigen wenigen Standard Templates einen Grossteil der für die Analyse einer klinischen Studie nötigen Tabellen relativ einfach und schnell zu erstellen, bildet den Kern des Konzepts, in dem PH-Clinical bei Schwarz Biosciences zum Einsatz kommt.

## 2 Das Konzept

Ausgangsbasis der statistischen Analyse sind SAS Datensätze, die aus der Oracle Clinical Datenbank exportiert werden. Aus technischen Gründen ist es bei Schwarz Biosciences nicht möglich, direkt auf den Oracle Daten zu arbeiten, so dass der „Umweg“ über den Export von SAS Datensätzen gegangen werden muss. Die SAS Datensätze, die aus Oracle exportiert werden, stellen eine 1:1 Abbildung der Oracle Datenbank dar und enthalten daher auch noch einige Variablen, die nur für die Dateneingabe erforderlich, aber für die Analyse überflüssig sind. Außerdem sind diese Daten Rohdaten, d.h. sie bilden den CRF ab und sind noch nicht für die Analyse geeignet, da noch abzuleitende Variablen fehlen und die Struktur für die Dateneingabe optimiert wurde und nicht für die Analyse. Es müssen also noch einige Änderungen an diesen Daten durchgeführt werden. Die fertigen Datensätze heißen dann Analyse-Datensätze.

Sobald die ersten Daten einer Studie vorliegen, werden die entsprechenden SAS Datensätze von Oracle erzeugt und danach in einer automatisierten Routine immer dann wieder, wenn neue Daten hinzu kommen. Diese ersten Rohdaten werden von überflüssigen Variablen bereinigt und in ihrer Struktur soweit angepasst, dass ein Upload nach PH-Clinical möglich ist. Unter PH-Clinical stehen nun die Rohdaten zum Review bereit und werden, wie oben beschrieben, bei Bedarf durch die aktuellen Datensätze ersetzt. Nun wird angefangen, die Modifikations-Programme zu schreiben, welche die Analyse Datensätze erzeugen. Dies sind im Prinzip ganz normale Base SAS Programme, die aber unter PH-Clinical als Template laufen, um die Vorteile des Audit Trails zu nutzen. Wenn die Datenbank vollständig ist und die finale Version aus Oracle exportiert wurde, laufen diese Programme und erzeugen die Analyse Datensätze. Diese werden dann nach PH-Clinical geladen und stehen nun für die Analyse zur Verfügung. Ein Hauptmerkmal sowohl der Analyse-Datensätze wie auch der Rohdaten ist, dass sie in den studienunabhängigen Modulen (z.B. Past Diseases, Past Medication, Demographics, Labordaten, Adverse Events, etc) immer die gleiche Struktur aufweisen und nur die Inhalte von Studie zu Studie variieren. Dieses soll auch bei studienabhängigen Modulen so weit wie möglich eingehalten werden, um das spätere Poolen von mehreren Studien zu vereinfachen.



**Abbildung 1:** Schematische Darstellung des Datenfluss

Zur Erstellung der statistischen Tabellen kann dann auf eine Template-Sammlung zurückgegriffen werden, mit der Standardtabellen, wie z.B. Summary Statistics und Häufigkeitstabellen einfach und schnell erstellt werden können. Da es unter PH-Clinical möglich ist, eine bestimmte Belegung der Platzhalter als sog. „report selection“ abzuspeichern, ist hier sogar eine weitere Vereinfachung möglich: Da studienunabhängige Variablen bei uns immer gleich heißen, wird eine Tabelle nur einmal erzeugt, die Belegung abgespeichert und anschließend kann diese Belegung für andere Studien immer wieder verwendet werden. Solche mehrfachen Abläufe können in einem sog. Batch-Book gespeichert werden

und dann auf einmal ablaufen. Dieser Prozess ist mit dem „Batch submit“ von Base SAS vergleichbar, nur das beliebig viele Templates mit beliebig vielen abgespeicherten Selektionen hintereinander laufen können.

### 3 Tipps und Tricks beim Programmieren

Im Prinzip reichen normale SAS Kenntnisse, um auch mit PH-Clinical zu arbeiten. Es sollten allerdings Kenntnisse im Bereich der SAS Macros vorhanden sein. Ausserdem ist zu beachten, dass das Template selber sehr abstrakt programmiert und daher auch nicht einfach auf Korrektheit zu überprüfen ist, denn vieles, was sonst konkret in einem Base SAS Programm steht und in einzelnen Schritten auf Korrektheit überprüft werden kann, wird hier per Platzhalter gekennzeichnet und erst zur Laufzeit mit „Leben“ gefüllt. PH Clinical Templates können auch nur im ganzen aufgerufen werden und nicht teilweise laufen. Als Programmierer muss man sich an viele sonst nicht so geläufige abstraktere Statements gewöhnen (call execute, oder verschachtelte Macro Variablen). Im Prinzip kann aber jeder versierte SAS Programmierer schnell in PH Clinical Templates programmieren.

### 4 Pro und Contra

PH Clinical hat viele Vorteile, die so nicht oder nur teilweise in Base SAS zur Verfügung stehen.

- Kontrollierter User Access
- Version Control für die Templates
- Audit Trail für alle PH-Clinical Objekte
- Zusammenfassung der Daten zu Studien
- Browsing-Tool zum Quality Check
- Volle Base SAS Funktionalität
- Universelle, flexible Templates, die benutzerfreundlicher sind als herkömmliche Macros.

Dem gegenüber steht eine sehr hohe Lizenzgebühr, der Zwang zur Sorgfalt bei der Erzeugung der Datensätze, da diese einmal nach PH-Clinical geladen nur durch erneuten upload modifizierbar sind und der Zwang zur Normierung der Daten (was nicht notwendigerweise ein Nachteil sein muss), um das volle Potential von PH-Clinical zu nutzen. Außerdem ist PH-Clinical nicht für ein Umfeld geeignet, in dem häufig wechselnde Datenstrukturen verwendet werden, da dann der Aufwand zur Harmonisierung der Daten den zu erwartenden Nutzen von PH-Clinical übersteigt.

