

Neuer Ansatz zur Einheiten-Konvertierung von Labordaten bei klinischen Studien

Steffen Witte

PRA GmbH & Universität Heidelberg
eMail: witte@imbi.uni-heidelberg.de

Abstract

Die Labordaten bei klinischen Studien sind häufig sehr grosse Dateien, 10000 oder auch 50000 Beobachtungen sind keine Seltenheit. Zur Vergleichbarkeit ist es nun oft erwünscht, wenn alle Werte eines Parameters (etwa Hämoglobin) in der gleichen Einheit vorliegen (etwa g/dl). Da bei multizentrischen Studien die Daten aus verschiedenen Quellen kommen, liegen in der Regel auch Werte in unterschiedlichen Einheiten vor. Eine Konvertierung ist hier nötig. Eine weitere Notwendigkeit zur Konvertierung ist der Vergleich der Messwerte mit gegebenen Normalwerten, die aber auch in anderen Einheiten gegeben sein können. Erst nach einer Konvertierung liegt eine Vergleichbarkeit vor und es kann entschieden werden, ob der individuelle Wert "out of range" ist oder nicht.

Die Umrechnung der Labordaten von gegebenen in gewünschte Einheiten kann je nach Menge und Vielfalt der Einheiten-Kombinationen zur Konfusion und viel Arbeit führen. Eine Automation ist also sinnvoll.

Unter einigen Prämissen soll in diesem Vortrag ein SAS[®] Macro (SAS[®] Base) zur einfachen und übersichtlichen Konvertierung von Labordaten vorgestellt werden. Im Kern geht es um die Erstellung einer Datei (mit Hilfe von proc sql), die alle Umrechnungsfaktoren enthält, so dass ein einfacher merge-Schritt ausreicht, um die Umrechnung auszuführen.

Einführung

Warum hat man verschiedene Einheiten für denselben Labor-Parameter innerhalb einer Studie?

Bei klinischen Multicenter Studien mit der Verwendung von lokalen Laboratorien (z.B. für jedes Zentrum ein Labor) hat üblicherweise jedes Labor einen eigenen Satz von Einheiten, die verwendet werden. Ebenso kann ein Wechsel von Laboratorien innerhalb einer Studie (insbesondere bei long term studys) verschiedene Einheiten für einen Labordatensatz mit sich bringen. Weiterhin können bei der Zusammenfassung von Labordatensätzen von verschiedenen Studien verschiedene Units auftauchen (integrierte DB / Meta-Analyse).

Wozu benötigt man eine einheitliche Einheit für einen Labor-Parameter?

Einheitlichkeit diesbezüglich ist natürlich nicht zwingend, zum Teil sogar nicht erwünscht. Häufig wird nur die Anzahl von Laborwerten ausgewertet, die über einer oberen Schranke oder unter einer unteren Schranke liegen (out of range analysis). Dafür werden die Normalwerte (Schranken) der jeweilig involvierten Laboratorien verwendet.

Oft ist es aber erwünscht, einen einheitlichen Satz von Normalwerten zu verwenden. Dafür müssen dann alle Werte eines Parameters in einer bestimmten Einheit vorliegen. Der zweite häufige Grund besteht darin, Statistiken aus den Laborwerten zu berechnen, und da nicht Äpfel mit Birnen verglichen werden sollen ...

Drei Typen von Konvertierungen

Um einen Wert, der in einer bestimmten Unit gemessen wurde, in eine andere Unit umzurechnen, benötigt man einen Umrechnungsfaktor (conversion factor). Es gibt drei verschiedene Typen von Faktoren.

Type I Konvertierung

Die Type I Konvertierung ist die wohl einfachste und auch häufigste Art der Umrechnung. Hierbei hängt der Konvertierungsfaktor nicht vom entsprechenden Labor-Parameter ab, sondern ausschliesslich von den Einheiten.

Zum Beispiel: Um von mg/dl in g/dl umzurechnen, verwendet man immer den Faktor 0.001:

Hämoglobin: 9500 mg/dl = 9.5 g/dl
Albumin: 3570 mg/dl = 3.57 g/dl

Type II Konvertierung

Bei einigen Einheiten spielt das spezifische Gewicht des entsprechenden Parameters eine Rolle. Muss man nun etwa von einer solchen Einheit in eine Einheit umrechnen, die nicht davon abhängig ist, wird wohl auch der Umrechnungsfaktor von dem entsprechenden Parameter abhängig sein.

Zum Beispiel: Um von g/dl in $\mu\text{mol/l}$ umzurechnen, sind verschiedene Faktoren nötig.

Hämoglobin: 1 g/dl = 620.7 $\mu\text{mol/l}$
Albumin: 1 g/dl = 144.9 $\mu\text{mol/l}$

Type III Konvertierung

Der nun seltenste Fall ist, dass die Umrechnung von dem Wert eines anderen Parameters abhängig ist. Das kommt insbesondere beim Differentialblutbild vor, bei dem unter Umständen die Typen der weissen Blutkörperchen (WBC) relativ zu der Gesamtanzahl von WBC angegeben werden oder werden sollen.

Zum Beispiel: Um von G/l in % umzurechnen, muss man einen individuellen Faktor ermitteln.

WBC: 6 G/l (= 6000 weisse Blutkörperchen / mm^3)
Eosinophils: 0.3 G/l = 5 %

Umsetzung in SAS[®]

1. Es soll jede Einheit nur einmal erfasst werden (und nicht alle Kombinationen von möglichen Umrechnungswünschen) und zwar mit der Relation zu einer "Basis-Einheit".
2. Aus diesem Datensatz kann anschliessend automatisch mit proc sql ein weiterer Datensatz erstellt werden, der alle möglichen Umrechnungskombinationen umfasst und der weiterhin den entsprechenden Umrechnungsfaktor bereitstellt.
3. Nun braucht der Original-Labor-Datensatz nur noch mit dem eben erstellten verknüpft zu werden (merge).
4. Um nun auch noch die Type III Konvertierung durchzuführen, bedarf es einiger Unterschritte, die ebenso angerissen werden sollen.

Schritt 1: Datensatz für die Units und Relationen

In diesem Datensatz, der die Grundlage darstellt, ist für die Gewichtseinheiten g/l (Gramm pro Liter) die "Basis-Einheit" und für die Einheiten, die Anzahlen pro Hohlmaß messen, ist es /l (Anzahl pro Liter). Die Units (UNITC) werden ausserdem der späteren Einfachheit halber zusätzlich mit einem numerischen Code (UNITN) versehen.

GROUP	UNITN	UNITC	TEST	RELATION
1	017	g/l	ALL	1
1	016	g/dl	ALL	10
1	012	mg/dl	ALL	0.01
1	022	mmol/l	CALB	69
1	022	mmol/l	HHB	16.11
2	135	/l	ALL	1
2	157	x10E3/l	ALL	1000
2	096	G/l	ALL	10E9
...

Die Variable GROUP verschlüsselt die Gruppe der Einheiten, die ineinander überführt werden können. In TEST steht der entsprechende Parameter, kann die Einheit in dem Eintrag aber generell (Type I Konvertierung) in die "Basis-Einheit" umgerechnet werden, so wird das Schlüsselwort ALL verwendet. Die RELATION gibt an, wie ein Wert in die "Basis-Einheit" umgerechnet werden kann, hier zum Beispiel ist der Faktor von g/dl nach g/l: 10, mit anderen Worten: $x \text{ g/dl} = x * 10 \text{ g/l}$. Nun ist Schritt eins bereits getan, zumindest für eine kleine Anzahl von Einheiten.

Schritt 2: Erstellung von Umrechnungskombinationen und den Faktoren

Nachdem der obigen Datensatz vergrößert wurde, damit für jeden Parameter, bei dem zuvor ALL stand, eine Zeile entsteht, kann man mit einem proc sql statement mit einem einfachen full join (kartesisches Produkt) den gewünschten Datensatz erstellen.

```
proc sql;
  create table conv_lis as

  select
    a.relation  as f1,
    b.relation  as f2,
    a.unitn    as u1,
    b.unitn    as u2,
    a.test     as param

  from
    conv_tab  as a,
    conv_tab  as b

  having a.group = b.group and param = b.test;

quit;
```

Anschliessend ist nur noch der Faktor zu berechnen, den man mit einem einfachen Dreisatz ermitteln kann. Der SAS[®] Code dazu beschränkt sich auf:

```
data conv_lis(keep = param from_to factor);
  set conv_lis;
  attrib
    from_to label = 'the unit combination for this conversion'
    factor label = 'the conversion factor';
  factor = f1 / f2;
  from_to = u1 + (u2 / 1000);
run;
```

Nun haben wir das folgende Ergebnis erhalten (Auszug). Dieser Datensatz kann immer wieder verwendet werden und ist je nach Wunsch einfach erweiterbar.

FROM_TO	PARAM	FACTOR
12.016	HHB	0.001
16.016	HHB	1
17.016	HHB	0.1
22.016	HHB	1.611
16.022	HHB	0.6207
96.135	HWBC	10E6
96.157	HWBC	10E9
96.096	HWBC	1
...		

Schritt 3: Merge: Faktoren und Labor Datensatz

Für das weitere Vorgehen werden einige Voraussetzungen verlangt:

- vertikale Datenstruktur, d.h. eine Observation für einen Wert
- Verwendung des numerischen 3stelligen Unit-Codes
- Verwendung des alphabetischen Parameter-Codes (HHB = Hämoglobin)
- Obiger Datensatz mit den Umrechnungsfaktoren wurde bereits erstellt
- Verwendung nur von SAS[®] Base nötig

Für das merging kann zum Beispiel ein Macro verwendet werden, um die Übersichtlichkeit im Programm zu wahren (%labconv).

Schritt 4: Type III Konvertierung

Die Konvertierung, bei der der Umrechnungsfaktor von einem anderen Wert des gleichen Datensatzes abhängt, kann nicht nur durch einen einfachen merge Schritt getan werden. Es soll hier nur eine mögliche Idee gegeben werden (%labconv3). Schritt 2 lässt erkennen, wie hilfreich ein Macro für die Type I und II Konvertierung ist.

1. Erzeuge eine Variable, die den Wert des Basis-Parameters enthält, ebenso eine weitere, die deren Einheit aufnimmt. Nun hat man alle Informationen in einer observation, die für die Umrechnung nötig ist.
2. Verwende nun eine Type I oder II Konvertierung, um die Einheit des Basis-Parameters in diejenige des zu konvertierenden Parameters umzurechnen.
3. Anschliessend können die Werte miteinander verglichen werden und die Umrechnung selbst - etwa in oder von % - kann vorgenommen werden (Dreisatz).

Beispiel

```
data labdata
```

```
input param2 $3-10 old oldunit newunit pt;
```

```
cards;
```

```
HHB    9    68   22   1
```

```
HWBC   6    96  140   1
```

```
HEOS   0.3  63   1    1
```

```
...
```

```
run;
```

```
%labconv (inset = labdata, newunit = newunit, oldvalue = old, param = param2)
```

(1 = %, 22 = mmol/l, 63 = /nl, 68 = g/100ml, 96 = G/L, 140 = / μ l)

Das würde das folgende Ergebnis liefern:

PARAM	OLD	OLDUNIT	NEWUNIT	PT	FACTOR	VALUE
HHB	9	g/100ml	mmol/l	1	0.62	5.59
HWBC	6	G/l	/ μ l	1	1000	6000
HEOS	0.3	/nl	%	1	.	.
...						

```
%labconv (inset = labdata, newunit = newunit, oldvalue = old, param = param2, varlist = HMONO HEOS, key = pt, basetest = HWBC);
```

Das würde das folgende endgültige Ergebnis liefern:

PARAM	OLD	OLDUNIT	NEWUNIT	PT	FACTOR	VALUE
HHB	9	g/100ml	mmol/l	1	0.62	5.59
HWBC	6	G/l	/ μ l	1	1000	6000
HEOS	0.3	/nl	%	1	16.67	5
...						