

## Validierung der Softwareinstallation von SAS 9.4

Friedrich-Wilhelm Röhl Institut für Biometrie und Medizinische Informatik der Otto-von-Guericke- Universität Leipziger Str. 44 39120 Magdeburg friedrich-wilhelm.roehl@med.ovgu.de	Siegfried Kropf Institut für Biometrie und Medizinische Informatik der Otto-von-Guericke- Universität Leipziger Str. 44 39120 Magdeburg siedfried.kropf@med.ovgu.de
---	--

### Zusammenfassung

Bei Programmsystemen, die ursprünglich in englischer Sprache entwickelt wurden und dann ins Deutsche übertragen wurden, stößt man ab und zu auf Anwendungsschritte, die im Deutschen nicht so zuverlässig funktionieren wie in der Originalsprache.

Das trifft im SAS auf das SAS Operational Qualification Tool zu, mit dem die korrekte Ausführung der installierten Prozeduren überprüft wird. Obwohl bei uns alle Einzelplatzlizenzen von der gleichen DVD installiert wurden, zeigt die englischsprachige Version eine korrekte Installation an, während mit der deutschen Ausführung 9 fehlerhafte Programmteile identifiziert wurden.

Die Ursache findet sich im Vergleich von Variablen aus den Rechnerergebnissen mit deutscher Bezeichnung, die nicht mit den englischen Referenzvariablen übereinstimmen. Die Rechenergebnisse selbst stimmen überein und liefern keinen Grund für eine Fehlermeldung.

**Schlüsselwörter:** Softwarevalidierung, SAS/STAT, SAS Qualification Tools

## 1 Einleitung

Softwareprodukte sind mittlerweile sehr komplex und umfangreich. Darüber hinaus wird die Installation der Software in der Regel automatisch gesteuert. Für den Anwender ist nur schwer zu erkennen bzw. zu überprüfen, ob die Software korrekt installiert wurde und fehlerfrei arbeitet.

Die Durchführung GCP-konformer klinischer Studien verlangt eine sorgfältige Programmierung der Analysen und eine Absicherung für korrekt erzeugte Ergebnisse [1]. Bevor man sich um die eigenen Programme kümmern kann, muss die verwendete Software validiert werden. SAS bietet für diesen Zweck zwei entsprechende Tools an:

- SAS Install Qualification Tool (SAS IQ) und
- SAS Operational Qualification Tool (SAS OQ),

mit denen überprüft werden kann, ob die SAS-Installation korrekt erfolgt ist.

Im Install Qualifikation Tool werden die Checksummen von Programmbestandteilen überprüft und im Operational Qualifikation Tool wird mit der auf dem Rechner instal-

lierten SAS-Version überprüft, ob die Ergebnisse der Prozeduren mit vorgegebenen Musterlösungen übereinstimmen.

## 2 Evaluationsprozess zur Überprüfung der SAS-Software für Einzelplatzlizenzen

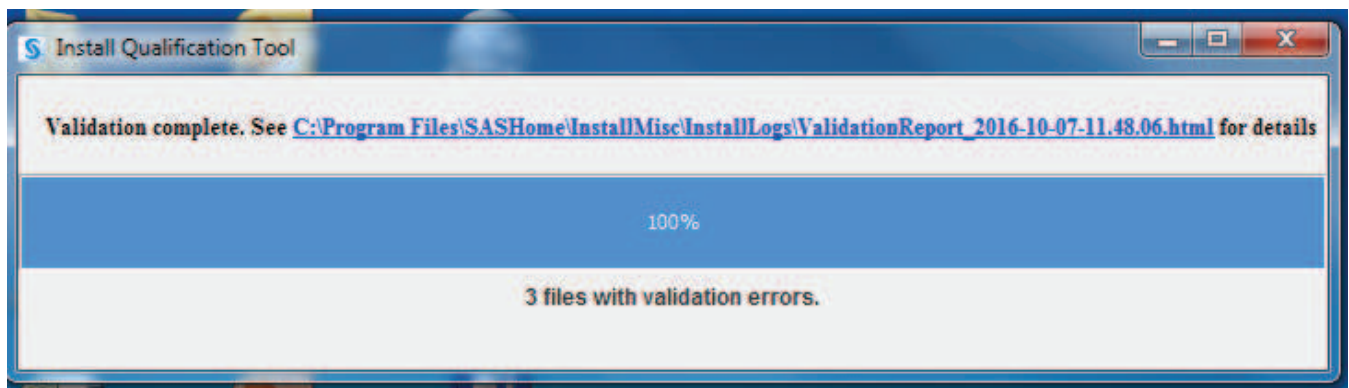
### 2.1 Installationsprozess

Wenn eine Einzelplatzinstallation mit den Standardeinstellungen durchgeführt wird, so findet man unter

Start → Alle Programme → SAS → UTILITIES →  
die beiden Programmbestandteile

- SAS Install Qualification Tool 9.4(65) und
- SAS Operational Qualification Tool

Das Install Qualifikation Tool überprüft bei unserer Einzelplatzlizenz die Checksummen von über 15000 Programmbestandteilen. Beim Aufruf startet das Programm ohne weitere Eingaben. Die Überprüfung dauert etwa 2 Minuten. Nach der initialen SAS-Installation sollten hier keine Fehler angezeigt werden. Wurde aber die Lizenz schon einmal erneuert oder nach der Erstinstallation neue Softwareprodukte hinzugefügt, meldet SAS Fehler bei der Überprüfung. Bei uns erscheint folgende Fehlermeldung:



**Abbildung 1:** Ergebnis der Validierung der SAS-Installation auf einem PC

Nach der jährlichen Lizenzerneuerung gibt es erwartungsgemäß Abweichungen, die in der zugehörigen SAS-Anleitung [2] bzw. im „ValidationReport“ beschrieben werden.

*When the SAS Installation Qualification Tool is run, the resulting validation is based on the state of the system at the time the initial installation completed. The tool's validation detects most manual modifications that the user makes to the SAS software components after the initial installation. So, for example, when the user installs software after an initial installation —such as adding additional products to the install or applying maintenance— the tool should not flag the additional files as having changed.*

Vergleichen wir eine Überprüfung vom Oktober 2016 mit einer Wiederholung im Februar 2017, erhalten wir unterschiedliche Checksummen mit nachfolgenden Hinweisen.

### Excluded Files

Excluded files are those for which the tool expected to find a difference. No action is required on the part of the user for these files.

```
[1]
SOFTWARE CODE:          base
FILE:
\SASFoundation\9.4\core\sashelp\core.sas7bcat
```

#### Checksum Data Does Not Match:

```
[1]
SOFTWARE CODE:          base
FILE:
\SASFoundation\9.4\core\sasinst\setinit.sss
PRE-INSTALL CHECKSUM:   94b4e6fa6bd6fd12e1990c240c8183bf
POST-INSTALL CHECKSUM:  179bdc08215f22f84120fa0243a1172d
```

```
[2]
SOFTWARE CODE:          base
FILE:
\SASFoundation\9.4\nls\de\sashelp\core.sas7bcat
PRE-INSTALL CHECKSUM:   c7c07e36164dfcbcb2d76bb950df4caa
POST-INSTALL CHECKSUM:  ec400c6286fef3d4f052d6ad9c29add7
```

Die erste Fehlermeldung war zu erwarten, genau wie die erste festgestellte Differenz der Checksumme für **setinit.sss**.

Die Validierung vom Februar 2017 am gleichen Rechner liefert:

```
[1]
SOFTWARE CODE:          base
FILE:
\SASFoundation\9.4\core\sasinst\setinit.sss
PRE-INSTALL CHECKSUM:   94b4e6fa6bd6fd12e1990c240c8183bf
POST-INSTALL CHECKSUM:  933cc917dd145165226bf5b5d6af077e
```

```
[2]
SOFTWARE CODE:          base
FILE:
\SASFoundation\9.4\nls\de\sashelp\core.sas7bcat
PRE-INSTALL CHECKSUM:   c7c07e36164dfcbcb2d76bb950df4caa
POST-INSTALL CHECKSUM:  6de6dd8665f376933c45414155dae2b3
```

Die Checksummenänderung für die **setinit.sss** ergibt sich nach Ablauf unseres Lizenzzeitraums am 31. Januar und neu eingespieltem Lizenzcodes.

Warum erhalten wir aber die dritte Fehlermeldung?

Diese Checksummenänderung, die hier angezeigt wird, hat offensichtlich mit der Sprachinstallation zu tun: ..... \nls\de\....

Auf dem Rechner wird mit deutscher Sprache gearbeitet.

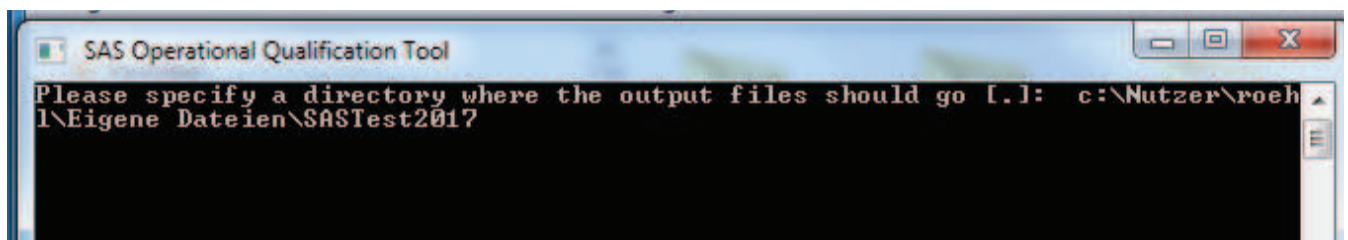
Da wir in der Sprache die Ursache für weitere Fehlermeldungen vermutet hatten, haben wir hier nach dem ersten Test manuelle Änderung vorgenommen, die von SAS aufgedeckt werden.

Darauf kommen wir in Verlaufe unserer weiteren Betrachtungen noch einmal zurück.

## 2.2 Überprüfung der Rechenergebnisse von Prozeduren

Bei der Anwendung des „SAS Operational Qualification Tool“ sollten keine Abweichungen auftreten, da hier die korrekte Arbeit der Prozeduren überprüft wird.

Nach Start dieses Evaluationstools muss in der Eingabeaufforderung zunächst ein Pfad angegeben werden, um zu definieren, wo das Protokoll der Überprüfung abgelegt wird:



**Abbildung 2:** Eingabemaske für den Ergebnispfad der Softwareüberprüfung

Nach der Überprüfung, die auf einem PC etwa 3-4 Minuten dauert, wird das Ergebnis in eine Datei mit dem Namen **ftt\_Datum.000** geschrieben, die die in Abbildung 3 dargestellten Unterverzeichnisse enthält.

Die Zusammenfassung der Überprüfung steht danach in der PDF-Datei **sasoq.pdf** bzw. im HTML-Dokument **sasoqtoc.htm**, deren erste Übersichtsseite in Abbildung 4 präsentiert wird.

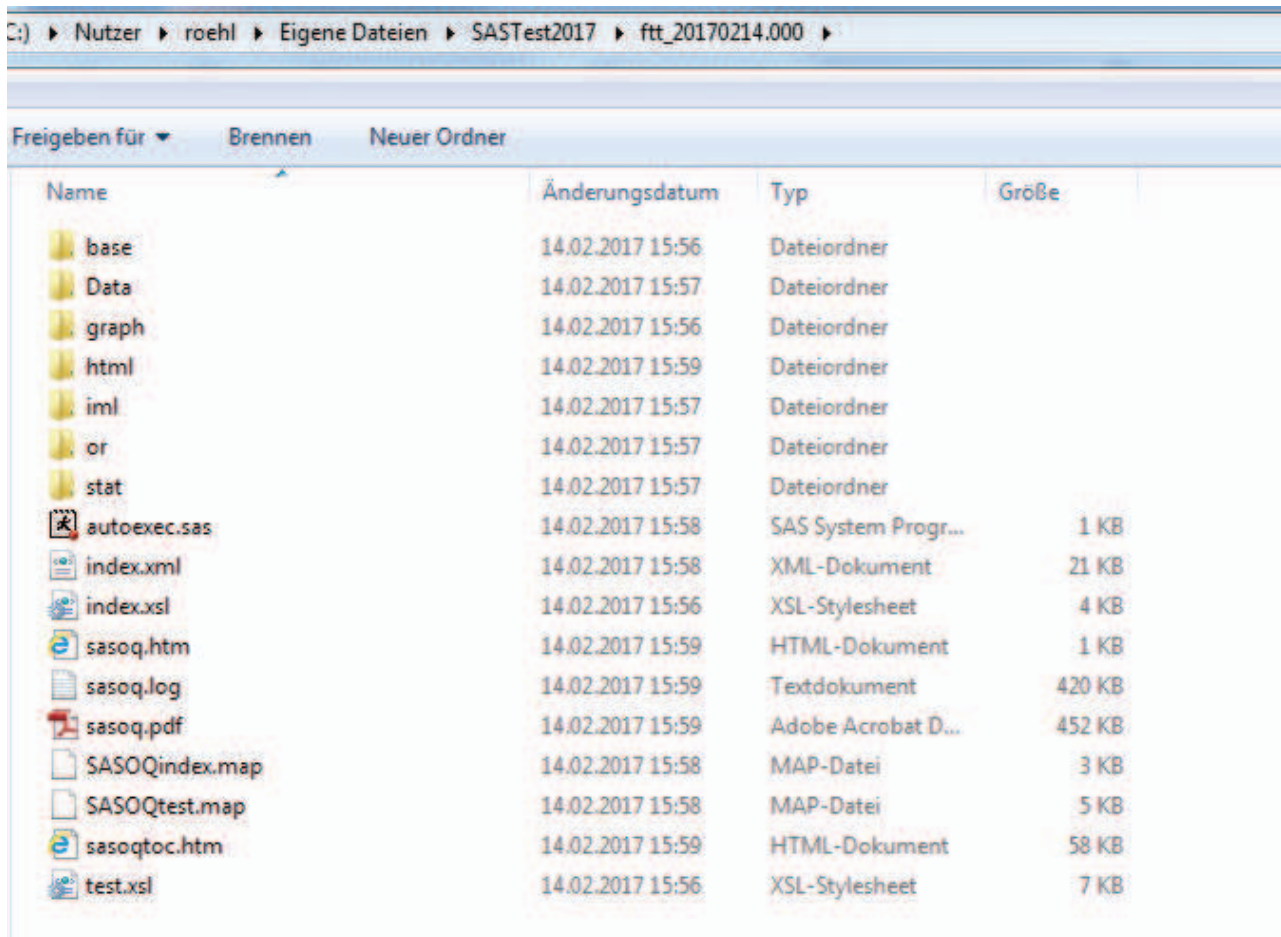


Abbildung 3: Dateiverzeichnis des Protokolls zu SAS OQ

Bei unserer Überprüfung zeigt das PDF-Dokument 9 fehlerhafte Programmbausteine:

### SAS Operational Qualification - Windows

Date:	2017-02-14
Time:	15:56:39
User:	roehl
Machine:	IBMI-81176
Platform:	Windows
SAS Version:	9.04.01M2P072314
Components Tested:	5
Tests Executed:	140
Tests Passed:	131
Tests Failed:	9

Abbildung 4: Zusammenfassung zur Validierung der Prozeduren in SAS/STAT

Konkret sind das die folgenden Bestandteile von SAS/STAT:

Nummer	Component	Table	Test
1	STAT	tststat	<b>tstglm02</b>
2	STAT	tststat	tstlft01
3	STAT	tststat	tstnst01
4	STAT	tststat	tstort01
5	STAT	tststat	tstpls01
6	STAT	tststat	tstprc01
7	STAT	tststat	tstrsr01
8	STAT	tststat	tststpo1
9	STAT	tststat	tstvcl01

Als Beispiel haben wir das Testprogramm **tstglm02** unter die Lupe genommen, um die Fehlerquelle zu ermitteln.

Im Protokoll zu diesem Programm steht folgenden Hinweis:

```
%check(pred,pred2,%str(method=absolute criterion=0.000001));
```

```
.....
```

```
278
```

```
279      data predinfo; set predinfo(keep=Statistic nValue1);
```

```
ERROR:  Die Variable Statistic in der DROP-, KEEP- oder RENAME-Liste wurde  
nie referenziert.
```

```
280
```

```
WARNING:  Datei WORK.PREDINFO wurde nicht ersetzt. Grund: da dieser Schritt  
angehalten wurde.
```

Um unsere Ergebnisse vorwegzunehmen: In der deutschen Version wird Statistik mit **k** geschrieben! Hier wird aber die englische Variable (Tabellenüberschrift) **Statistic** überprüft.

Für die Kontrolle haben wir das Programm **tstglm02.sas** aus

```
C:\Programme\SASHome\SASFoundation\9.4\sastest\stat\tstglm02.sas
```

separate benutzt und uns die Datenmatrix **PREDINFO** genauer angesehen, die aus folgendem Programmteil erzeugt wird:

```
data cov;
  input a b y z;
  cards;
1 1 74 3
1 1 68 4
1 1 77 5
1 2 69 4
1 2 74 6
1 2 71 5
```

```

2 1 76 2
2 1 80 4
2 2 79 4
2 2 76 5
3 1 87 3
3 1 91 7
3 2 91 6
3 2 77 4
;
proc glm data=cov;
  ods output ClassLevels          = levels
             NObs                 = nobs
             XPX                  = xpx
             InvXPX               = xpxi
             GEstFunc             = gestfunc
             EstFunc              = estfunc
             OverallANOVA        = anova
             FitStatistics        = stat
             ModelANOVA          = modanova
             ParameterEstimates  = est
             PredictedValues     = pred
             PredictedInfo       = predinfo
             Means               = means
             LSMeanCoef          = lsmcoef
             LSMeans             = lsmeans
             Diff                = diff;

  class a b;
  model y=a b z /solution xpx p i e e1 e2 e3 e4;
  means a;
  lsmeans a /e stderr pdiff;
run;

```

Die Ergebnisdatei **predinfo** enthält folgende Informationen:

Beob.	Dependent	Statistik	cValue1	nValue1
1	y	Residuensumme	0.0000000	5.684342E-14
2	y	Quadratsummen-Residuen	110.5515152	110.551515
3	y	Quadratsummen Residuen - Fehler SS	-0.0000000	-7.10543E-14
4	y	Autokorrelation erster Ordnung	-0.6609776	-0.660978
5	y	Durbin-Watson D	2.9793599	2.979360

Die deutsche Bezeichnung der dritten Merkmalspalte passt jetzt aber nicht mit dem folgenden Programmabschnitt aus dem Kontrollprogramm **tstglm02.sas** zusammen, das an der entsprechenden Stelle die englische Bezeichnung **Statistic** hat!

Dadurch wird auch der **keep**-Befehl nicht ausgeführt und die Variable **cValue1** verbleibt in der Datei und fehlt entsprechend in der Kontrollmatrix.

```
data predinfo; set predinfo(keep=Statistic nValue1);  
proc print data=predinfo;  
run;  
data predinfo2;  
    input nValue1;  
    cards;  
0.000000000  
110.5515152  
0.000000000  
-.660977601  
2.979359933  
;  
data predinfo(drop=statistic); set predinfo; run;
```

Im Log-Fenster gibt es deshalb folgenden Hinweis:

*NOTE: There were 5 observations read from the data set WORK.PREDINFO.*

*80 data predinfo; set predinfo(keep=Statistic nValue1);  
ERROR: Die Variable **Statistic** in der DROP-, KEEP- oder RENAME-Liste wurde  
nie referenziert.*

*NOTE: The SAS System stopped processing this step because of errors.*

*WARNING: The data set WORK.PREDINFO may be incomplete. When this step was  
stopped there were 0 observations and 0 variables.*

*WARNING: Datei WORK.PREDINFO wurde nicht ersetzt. Grund: da dieser Schritt  
angehalten wurde.*

Ersetzt man aber **StatistiC** durch die deutsche Bezeichnung **StatistiK**, läuft das Programm ohne Fehlermeldung durch. Soll hier auch die Checksumme überprüft werden, muss vorher das Programm **assert.sas** aus C:\Programme\SASHome\SASFoundation\9.4\sastest\base\assert.sas durchlaufen, um das Makro **rccheck** bereitzustellen, das die Rückgabecodes überprüft.

Eine Übersicht zu den Programmbestandteilen, die als Fehlerquelle für die übrigen 8 Programmteile identifiziert wurden, gibt die Tabelle im Anhang.

Den gleichen Effekt erzielt man, wenn man die Spracheinstellung in der config-Datei ändert. Wenn auf dem Rechner eine deutsche Version benutzt wird, findet man im Verzeichnis C:\SASHome\SASFoundation\9.4\sasv9.cfg eine config-Datei, die aus folgender Zeile besteht:

```
-config "C:\Program Files\SASHome\SASFoundation\9.4\nls\de\sasv9.cfg"
```

Ändert man hier die Spracheinstellung von deutsch (**de**) auf „englisch“ (**en**) gibt es beim Testdurchlauf keine Fehlermeldungen. Man arbeitet in diesem Moment mit der englischen Version.



Somit erhält man die Bestätigung, dass die Installation korrekt ausgeführt wurde und die Prozeduren richtige Ergebnisse liefern.

## Literatur

- [1] Gregory S. Nelson, Jay Zhou. Good Programming Practices in Healthcare Creating Robust Programs. Paper417-2012, SAS Global Forum 2012
- [2] SAS Institute Inc.2013. SAS 9.4 Qualification Tools User’s Guide, Cary, NC: SAS Institute Inc.

## Anhang

**Tabelle 1:** Übersicht über die von uns festgestellten Differenzen zwischen deutscher und englischer SAS-Installation, die zu Problemen bei der Softwarevalidierung führen

Testprogramm	Datenmatrix	deutsch	englisch	Ursache (Kontrollmatrix)
Tstglm02	predinfo	Statistik cValue1	statistic -	Schreibweise drop nicht ausgeführt
Tstlft01	Loghom Wilhom	Schicht Schicht	group group	Merkmalsbezeichnung
tstnst01	Statistics	Statistik	statistic	Schreibweise
tstort01	FitStatistics	Statistik	-	drop=statistic
tstpls01	Codedcoef	Effekt	effect	Schreibweise
tstprc01	NobsNVar	Beschreibung	-	drop=description
tstrsr01	Fitstatistics	Statistik	-	drop=statistic
tststpo1	BSSCP TSSCP	Etikett Etikett	- -	drop=label drop=label
tstvcl01		Etikett	-	drop=label