Validation and Compliance Issues

Helmut Schütz

Validated hardware?

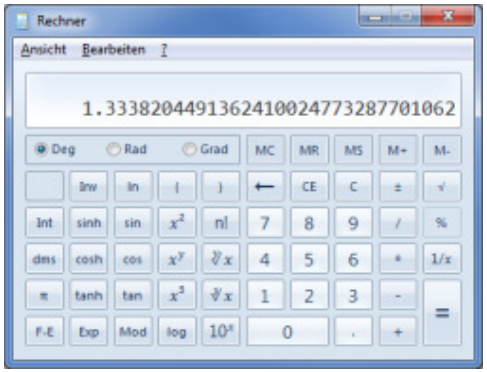
Pentium FDIV bug (1993).

- Flaw in the x86 assembly language floating point divison.
  - Example

$$\frac{4,195,835}{3,145,727} = 1.333739068902037589$$

$$\frac{4,195,835}{3,145,727} = 1.333820449136241002$$

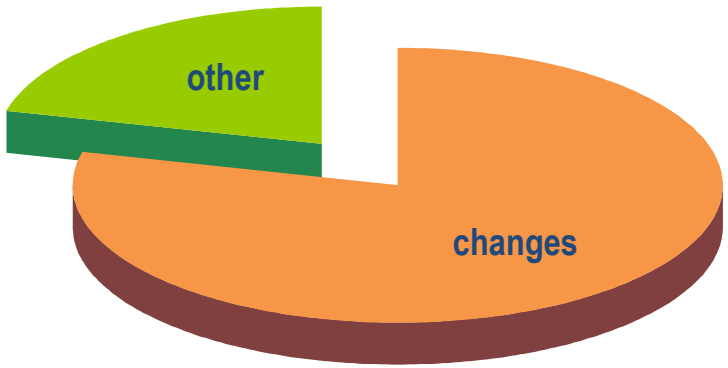
  - Costs for replacement: $475 million.



Validated software?

General Principles of Software Validation (FDA 2002).

- Section 2.4: Regulatory Requirements for Software Validation
  - 242 FDA Medical Device Recalls attributed to software failures (1992 – 1998).
  - 192 (79%) caused by software defects that were introduced when *changes* were made to the software after its initial pro-duction and distribution.



Spreadsheets?

Radio Yerevan Jokes.

- Radio Yerevan was asked:
  Is it possible to validate M$ Excel?
- Radio Yerevan anwered:
  In principle yes, but only if you buy the source code from Mr Gates first.

EMA CPMP/CHMP/EWP (Q&A 2011–2015)

- Results obtained by alternative, validated statistical programs are also acceptable except spreadsheets because outputs of spreadsheets are not suitable for secondary assessment.

Esch *et al.* 2010. *Good Laboratory Practice (GLP) – Guidelines for the Development and Validation of Spreadsheets.* doi:10.1002/qaj.466.

Spreadsheets?

M$ Article 828888: 'You can expect that for most users, such round off errors are not likely to be troubling in practice.'

MS Excel 1985 – 2002.

	A	B	C	D	E	F
1	0	formula (A)	100,000,000	formula (C)	1	formula (E)
2	–1	=A$1–1	99,999,999	=C$1–1	0.99999999	=E$1–0.00000001
3	±0	=A$1	100,000,000	=C$1	1.00000000	=E$1
4	+1	=A$1+1	100,000,001	=C$1+1	1.00000001	=E$1+0.00000001
5	1	=STDEV(A2:A4)	0	=STDEV(C2:C4)	0	=STDEV(E2:E4)

In calculating the 90% CI we need the *t*-distribution (for α 0.05 and the residual degrees of freedom).

- Example: $t_{0.05,\,22} = 1.717$.
- However, in MS Excel <2007:

	A	B	C	D	E	F	G	H
1	α	df	t	formula (C)	t	workaround (E)	t	Excel 2007+
2	0.05	22	2.074	=TINV(A2, B2)	1.717	=TINV(2*A2, B2)	1.717	=T.INV(A2, B2)

Open source software?

In principle yes – if it's validated, *why not*?

- Since the source code is accessible, even a 'white box' validation – which *no* (!) off-the-shelf software offers – is possible.
  - The FDA regularly uses R in M&S itself
    (but – as an agency – is not obliged to validate anything …).
  - New releases/updates more frequent than commercial SW.
    - R & packages: 3 – 4 / year.
    - Defects in packages: Generally corrected within one week.
  - R-packages relevant for bioequivalence:
    - Randomization: `randomizeBE` (2017)
    - NCA/BE: `bear` (2017)
    - Power and sample size: `PowerTOST` (2017)
    - Two-Stage Designs: `Power2Stage` (2017)

The R Foundation for Statistical Computing. Vienna; 2014: *R: Regulatory Compliance and Validation Issues.* http://www.r-project.org/doc/R-FDA.pdf.

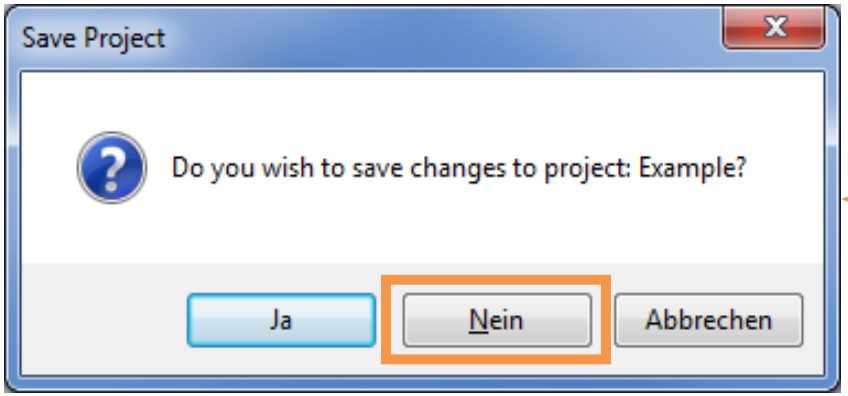
Alterations of data possible?

Example: Phoenix/WinNonlin

Properties | Information | History

Example >> Data >> EMA full replicate

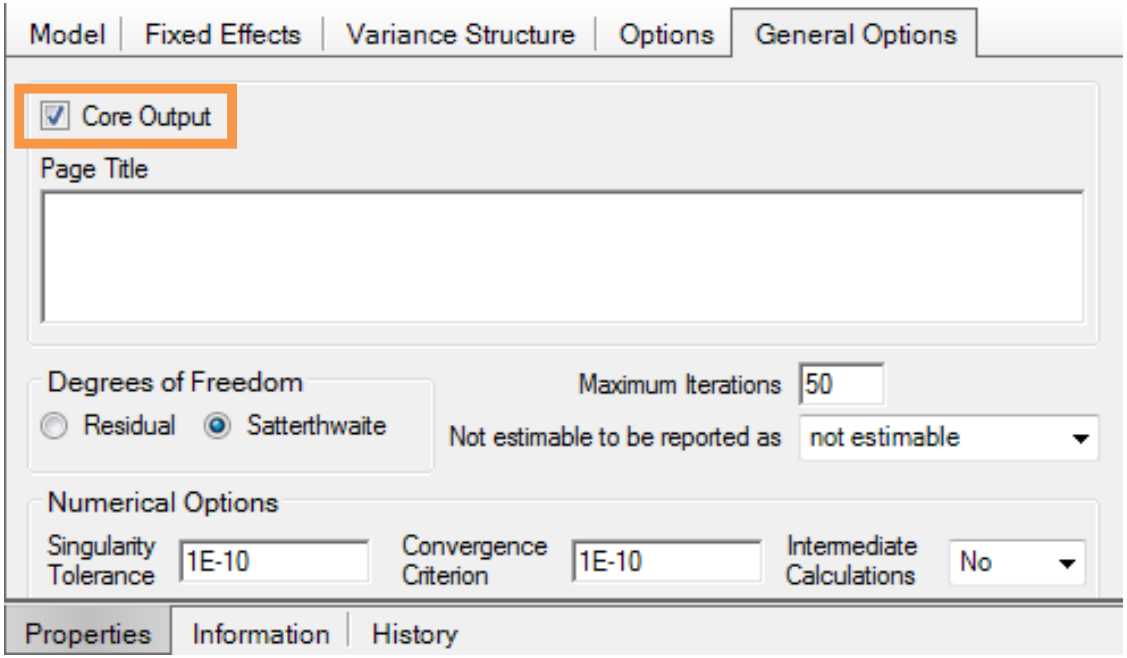
Timestamp	User	Object Name	Event	Description
2015.04.09 14:45:10 UTC	HS	Worksheet	Object Created	Object Created E:\Public\Documents\BEBAC\Phoenix Projects\EMA full replicate.xls
2015.04.09 14:45:46 UTC	HS	EMA full replicate	Value changed	F1; changed from 7.734541 to 7.5



If software allows changes without an audit trail, take measures!
PKS is 21 CFR 11 compliant…

Alterations of data possible?

Example: Phoenix/WinNonlin



Always select the Core Output (off by default)

Alterations of data possible?

Example: Phoenix/WinNonlin

```
                                                    Date:     4/09/2015
                                                    Time:     17:20:50

                WINNONLIN LINEAR MIXED EFFECTS MODELING / BIOEQUIVALENCE
                                        6.4.0.768
                                Core Version 30Jan2014


  Model Specification and User Settings
          Dependent variable : Data
                   Transform : LN
                Fixed terms  : int+Sequence+Subject(Sequence)+Period+Formulation
      Singularity tolerance  :       1e-010
      Denominator df option  : satterthwaite
```

Only in the Core Output you get a timestamp of the evaluation.
Avoid fancy Excel- or Word-Export options (if possible).

Old hats …

Parallel Groups: Example

- Evaluation (modified data set)

Program	equal variances	unequal variances
R 2.5.0 (2007)	81.21% – 190.41%	76.36% – 202.51%
NCSS 2001 (2001)	81.21% – 190.41%	76.36% – 202.51%

- Inflated α-risk in 'conventional' *t*-test (naive pooling) is reflected in a tighter confidence interval.
- Preliminary testing for equality in variances is flawed*) and should be avoided (FDA).
- Approximations (*e.g.*, Satterthwaite, Aspin-Welch, Howe, Milliken-Johnson) are currently *not implemented* in packages 'specialized' in BE (WinNonlin, Kinetica, EquivTest/PK)!

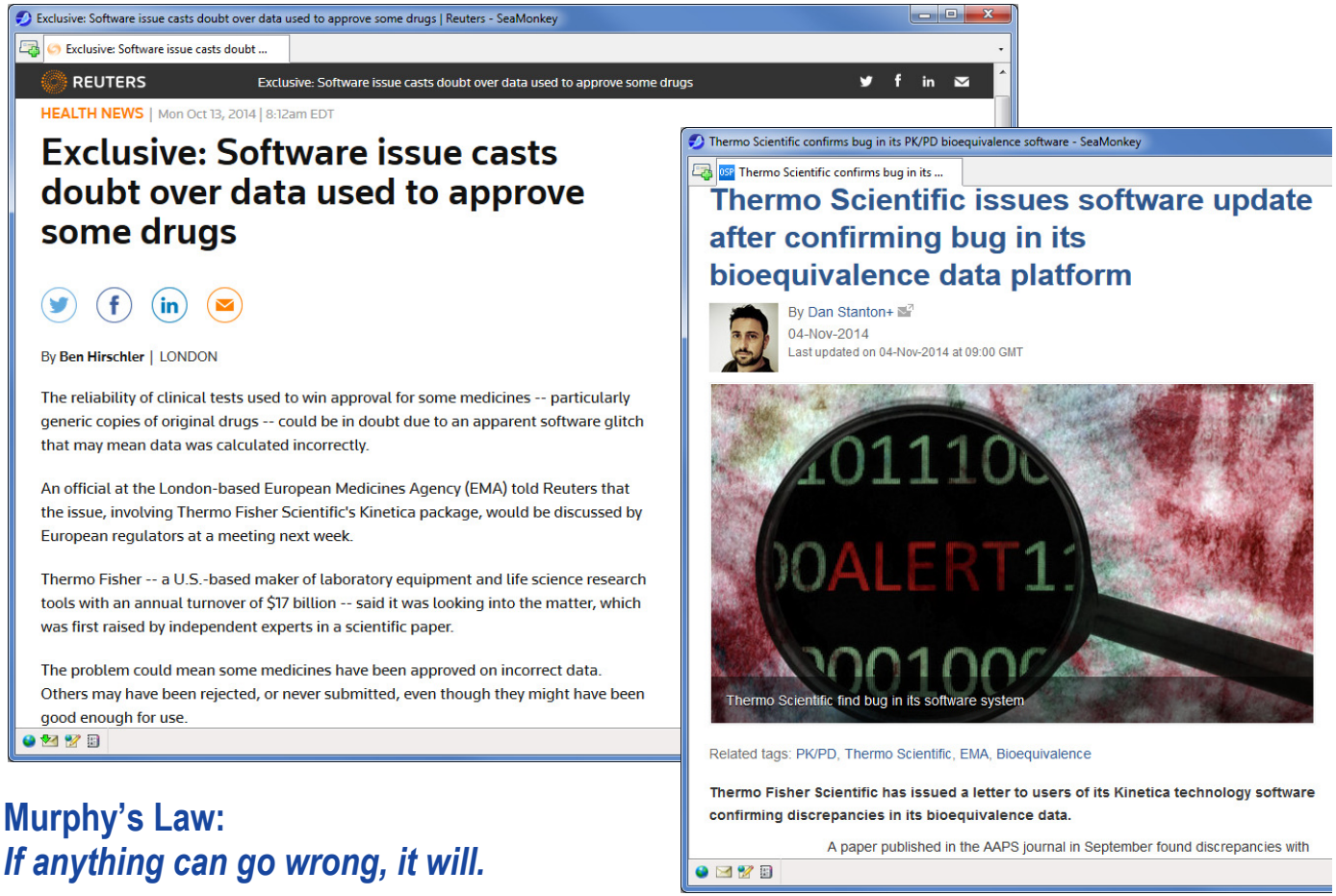
Surprise?

*) Moser, B.K. and Stevens, G.R.;
Homogeneity of variance in the two-sample means test.
Amer. Statist. 46, 19-21 (1992)

informa life sciences

Dissolution Testing, Bioavailability & Bioequivalence | Budapest, 24 May 2007 38

… making it to the health news



Exclusive: Software issue casts doubt over data used to approve some drugs | Reuters - SeaMonkey

Exclusive: Software issue casts doubt ...

REUTERS

Exclusive: Software issue casts doubt over data used to approve some drugs

HEALTH NEWS | Mon Oct 13, 2014 | 8:12am EDT

Exclusive: Software issue casts doubt over data used to approve some drugs

By **Ben Hirschler** | LONDON

The reliability of clinical tests used to win approval for some medicines -- particularly generic copies of original drugs -- could be in doubt due to an apparent software glitch that may mean data was calculated incorrectly.

An official at the London-based European Medicines Agency (EMA) told Reuters that the issue, involving Thermo Fisher Scientific's Kinetica package, would be discussed by European regulators at a meeting next week.

Thermo Fisher -- a U.S.-based maker of laboratory equipment and life science research tools with an annual turnover of $17 billion -- said it was looking into the matter, which was first raised by independent experts in a scientific paper.

The problem could mean some medicines have been approved on incorrect data. Others may have been rejected, or never submitted, even though they might have been good enough for use.

Thermo Scientific confirms bug in its PK/PD bioequivalence software - SeaMonkey

Thermo Scientific confirms bug in its ...

Thermo Scientific issues software update after confirming bug in its bioequivalence data platform

By Dan Stanton+

04-Nov-2014

Last updated on 04-Nov-2014 at 09:00 GMT

Thermo Scientific find bug in its software system

Related tags: PK/PD, Thermo Scientific, EMA, Bioequivalence

Thermo Fisher Scientific has issued a letter to users of its Kinetica technology software confirming discrepancies in its bioequivalence data.

A paper published in the AAPS journal in September found discrepancies with

Murphy's Law:
If anything can go wrong, it will.

Reference data sets in BE

Different software (general purpose, specialized in BE, commercial and open source), 2×2×2 cross-over.

DS	EquivTest		Kinetica		SAS		WinNonlin		R	
A	90.76	99.62	90.76	99.62	90.76	99.62	90.76	99.62	90.76	99.62
B	51.45	98.26	51.45	98.26	51.45	98.26	51.45	98.26	51.45	98.26
C	39.41	87.03	44.91	99.31	39.41	87.03	39.41	87.03	39.41	87.03
D	51.45	98.26	51.45	98.26	51.45	98.26	51.45	98.26	51.45	98.26
E	55.71	151.37	55.71	151.37	55.71	151.37	55.71	151.37	55.71	151.37
F	93.37	106.86	93.37	106.86	93.37	106.86	93.37	106.86	93.37	106.86
G	88.46	95.99	88.46	95.99	88.46	95.99	88.46	95.99	88.46	95.99
H	86.81	100.55	107.80	115.85	86.81	100.55	86.81	100.55	86.81	100.55

A, B, D – G Balanced sequences ($n_{TR} = n_{TR}$)
C, H Imbalanced sequences ($n_{TR} \neq n_{RT}$)

Schütz H, Labes D, Fuglsang A. 2014. *Reference Datasets for 2-Treatment, 2-Sequence, 2-Period Bioequivalence Studies.* doi:10.1208/s12248-014-9661-0.
Moralez-Acelay *et al.* 2015. *On the Incorrect Statistical Calculations of the Kinetica Software Package in Imbalanced Designs.* doi:10.1208/s12248-015-9749-1.

Reference data sets in BE

Two-group parallel (conventional *t*-test).

DS	EquivTest		Kinetica		SAS		WinNonlin		OO Calc		R	
1	27.15	86.94	27.15	86.94	27.15	86.94	27.15	86.94	27.15	86.94	27.15	86.94
2	18.26	96.59	15.76	119.00	18.26	96.59	18.26	96.59	18.26	96.59	18.26	96.59
3	26.35	415.71	26.35	415.71	26.35	415.71	26.35	415.71	26.35	415.71	26.35	415.71
4	38.60	134.21	38.60	134.21	38.60	134.21	38.60	134.21	38.60	134.21	38.60	134.21
5	106.44	112.10	106.39	112.44	106.44	112.10	106.44	112.10	106.44	112.10	106.44	112.10
6	91.85	115.78	92.07	115.50	91.85	115.78	91.85	115.78	91.85	115.78	91.85	115.78
7	106.86	126.49	104.30	129.32	106.86	126.49	106.86	126.49	106.86	126.49	106.86	126.49
8	105.79	113.49	105.79	113.49	105.79	113.49	105.79	113.49	105.79	113.49	105.79	113.49
9	103.80	120.61	103.80	120.61	103.80	120.61	103.80	120.61	103.80	120.61	103.80	120.61
10	107.20	126.99	104.59	130.16	107.20	126.99	107.20	126.99	107.20	126.99	107.20	126.99
11	7.83	17.38	6.98	19.51	7.83	17.38	7.83	17.38	7.83	17.38	7.83	17.38

1, 3, 4, 8, 9 Equal group sizes ($n_T = n_R$)

2, 5 – 7, 10, 11 Unequal group sizes ($n_T \neq n_R$)

Fuglsang A, Schütz H, Labes D. 2015. *Reference Datasets for Bioequivalence Trials in a Two-Group Parallel Design.* doi:10.1208/s12248-014-9704-6.

Reference data sets in BE

Two-group parallel (Welch's test).

DS	SAS		WinNonlin*		OO Calc		R	
1	26.78	88.14	26.78	88.14	26.78	88.14	26.78	88.14
2	23.71	74.38	23.71	74.38	23.71	74.38	23.71	74.38
3	24.40	449.08	24.40	449.08	24.40	449.08	24.40	449.08
4	38.05	136.15	38.05	136.15	38.05	136.15	38.05	136.15
5	106.44	112.10	106.44	112.10	106.44	112.10	106.44	112.10
6	91.84	115.79	91.84	115.79	91.84	115.79	91.84	115.79
7	97.38	138.51	NA		97.38	138.51	97.38	138.51
8	105.79	113.49	NA		105.79	113.49	105.79	113.49
9	103.80	120.61	NA		103.80	120.61	103.80	120.61
10	97.82	139.17	NA		97.82	139.17	97.82	139.17
11	6.30	21.60	NA		6.30	21.60	6.30	21.60

1, 3, 4, 8, 9 Equal group sizes ($n_T = n_R$)

2, 5 – 7, 10, 11 Unequal group sizes ($n_T \neq n_R$)

* Workaround required in WinNonlin; limited to 1,000 subjects / group.

Welch's test not implemented in EquivTest and Kinetica.

Fuglsang A, Schütz H, Labes D. 2015. *Reference Datasets for Bioequivalence Trials in a Two-Group Parallel Design.* doi:10.1208/s12248-014-9704-6.

Likely cause of Kinetica's defects

2×2×2 cross-over

- Calculation of the confidence interval (CI):

$$CI = e^{\log(\bar{x}_T - \bar{x}_R) \pm t_{1-\alpha,\, n_{RT}+n_{TR}-2} \sqrt{\frac{MSE}{2}\left(\frac{1}{n_{TR}} + \frac{1}{n_{RT}}\right)}}$$

- *Only* if sequences are balanced ($n_{TR} = n_{TR}$) a simplified formula based on the total sample size *N* is correct:

$$CI = e^{\log(\bar{x}_T - \bar{x}_R) \pm t_{\alpha,\, n_{RT}+n_{TR}-2} \sqrt{\frac{2\,MSE}{N}}}$$

Likely cause of Kinetica's defects

Two-group parallel

- Calculation of the confidence interval (CI):

$$CI = e^{\log(\bar{x}_T - \bar{x}_R) \pm t_{1-\alpha,\, n_T+n_R-2} \sqrt{\frac{MSE}{2}\left(\frac{1}{n_T}+\frac{1}{n_R}\right)}}$$

- According to the manual Kinetica uses a 'simplified' formula – but the sample size of subjects receiving the reference [*sic*] treatment in the denominator:

$$CI = e^{\log(\bar{x}_T - \bar{x}_R) \pm t_{1-\alpha,\, n_T+n_R-2} \sqrt{\frac{2\,MSE}{n_R}}}$$

Validation and Compliance Issues

Thank You!
Open Questions?

Helmut Schütz

BEBAC

Consultancy Services for
Bioequivalence and Bioavailability Studies
1070 Vienna, Austria

helmut.schuetz@bebac.at