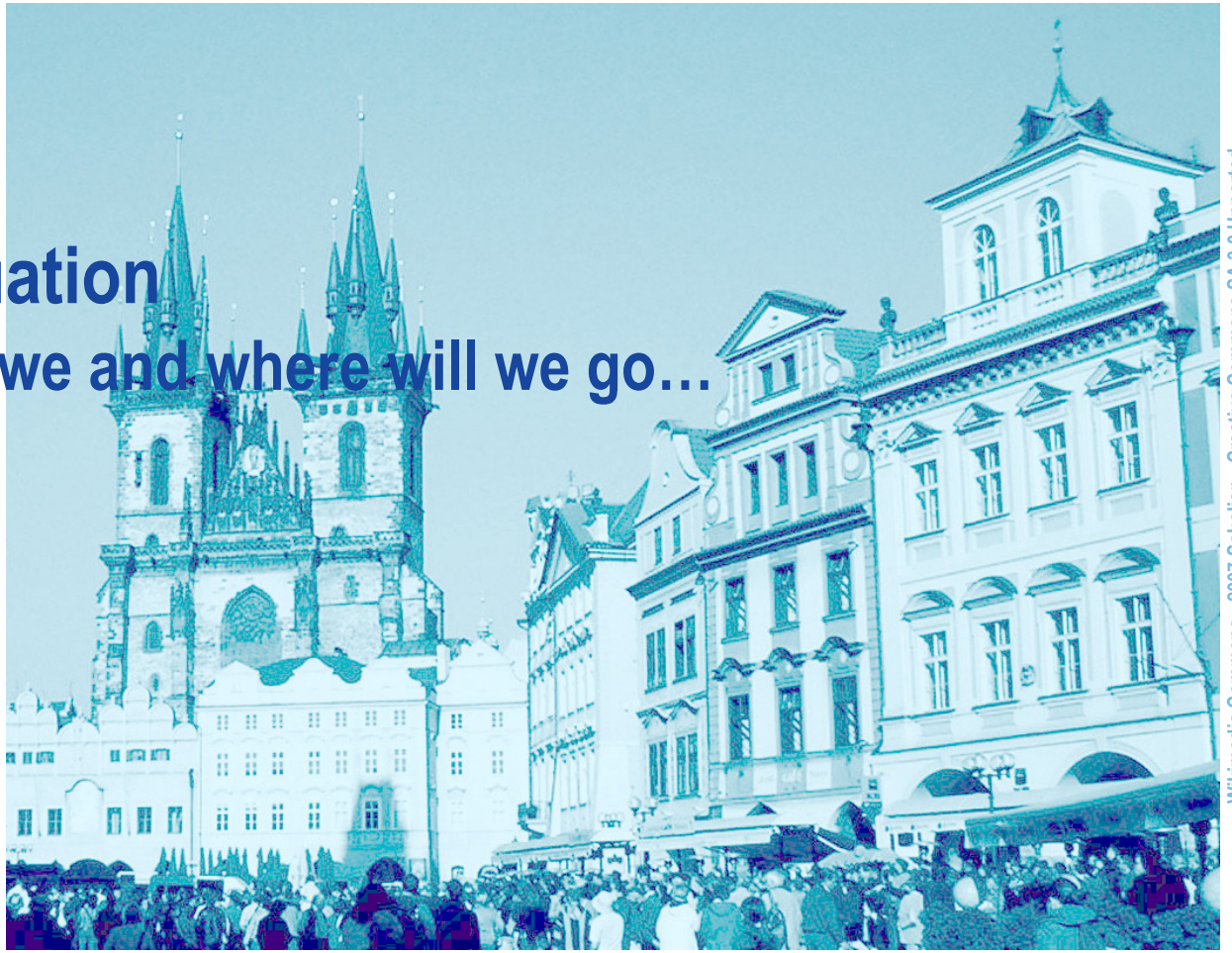
t_{max} Evaluation
Where are we and where will we go…

Helmut Schütz



Review of Guidelines

EC (1991), EMEA (2001)

Statistical evaluation of t_{max} only makes sense if there is a clinically relevant claim for rapid release or action or signs related to adverse effects. The non-parametric 90 % confidence interval for this measure of relative bioavailability should lie within a clinically determined range.

FDA, Health Canada (since 1992)

No comparison of t_{max}. If relevant, early partial *AUC*.

FDA: Cut-off time median t_{max} of reference

HC: Cut-off time subject's t_{max} of reference

Argentina, Japan, South Africa (current)

Only if clinically relevant, comparison of t_{max} by non-parametric statistical methods.

Review of Guidelines

EMA (BE GL 2010)

A statistical evaluation of t_{max} is not required. However, if rapid release is claimed to be clinically relevant and of importance for onset of action or is related to adverse events, there should be no apparent difference in median t_{max} and its variability between test and reference product.

- What might 'apparent' be?
- The median is a certain number – it does not have a 'variability'

EMA (MR GL 2014)

For delayed and multiphasic release formulations differences in t_{max} is also recommended to be assessed, especially for products where a fast onset of action is important. A formal statistical evaluation of t_{max} is not required. However, there should be no apparent difference in median t_{max} and its range between test and reference product.

- The range has a breakdown point of zero

Review of Guidelines

ASEAN states, Australia, Chile, Eurasian Economic Union, members of the Gulf Cooperation Council, New Zealand (current)

- The EMA's vague recommendation of 2010 incurred

WHO (2017)

- Where t_{max} is considered clinically relevant, median and range of t_{max} should be compared between test and comparator to exclude numerical differences with clinical importance. A formal statistical comparison is rarely necessary. Generally the sample size is not calculated to have enough statistical power for t_{max}. However, if t_{max} is to be subjected to a statistical analysis, this should be based on non-parametric methods and should be applied to untransformed data.

Review of Guidelines

EMA (draft product-specific guidances 2022)

Comparable median (≤ 20 % difference) and range for T_{max}.

In a footnote:

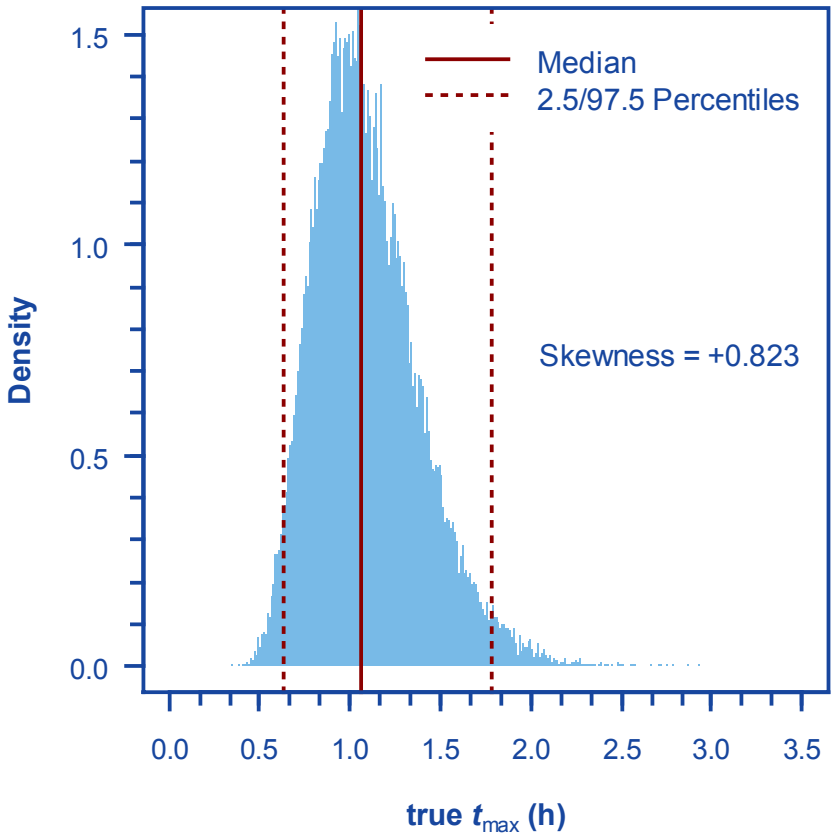
This revision concerns defining what is meant by 'comparable' T_{max} as an additional main pharmacokinetic variable in the bioequivalence assessment section of the guideline.

- **Still: What is a 'comparable' range?**

Statistical Properties of t_{max}

The *true* (but *unknown*) t_{max} follows a <u>continuous</u> distribution on a <u>ratio scale</u>

- Transformations
  Any suitable
- Allowed operations
  Difference, ratio



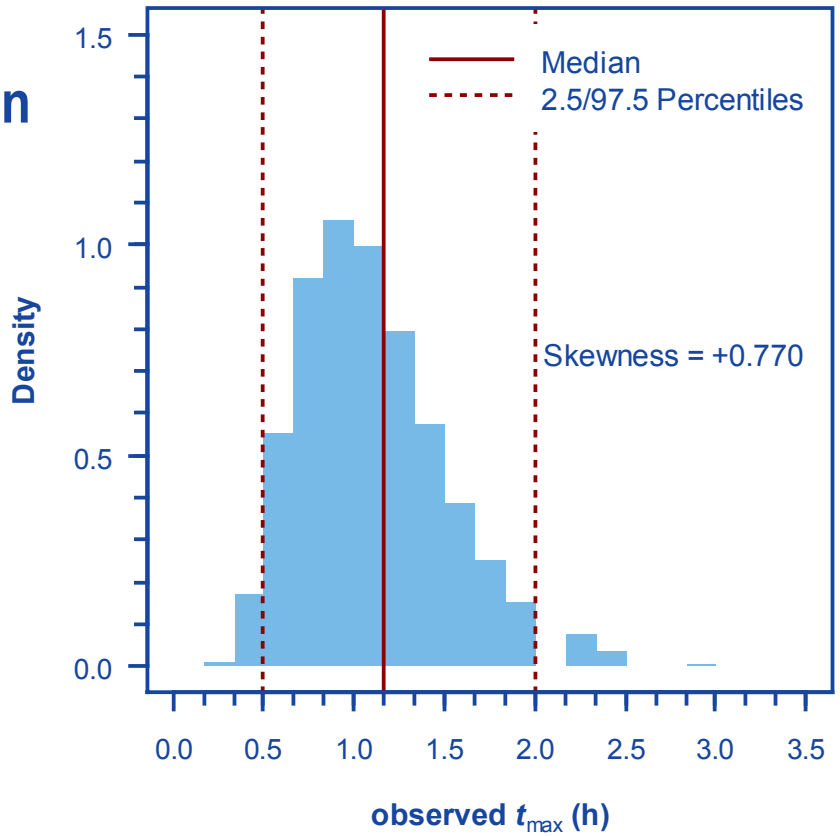
One-compartment model: k_{01} 3.037, k_{10} 0.1733 h^{-1} ($t_{½}$ 4 h), no lag-time, theoretical t_{max} 1 h; 50,000 simulated profiles

Statistical Properties of t_{max}

The *observed* t_{max} follows a discrete distribution on an ordinal scale

- Transformations
  None
- Only (!!) allowed operation
  Difference

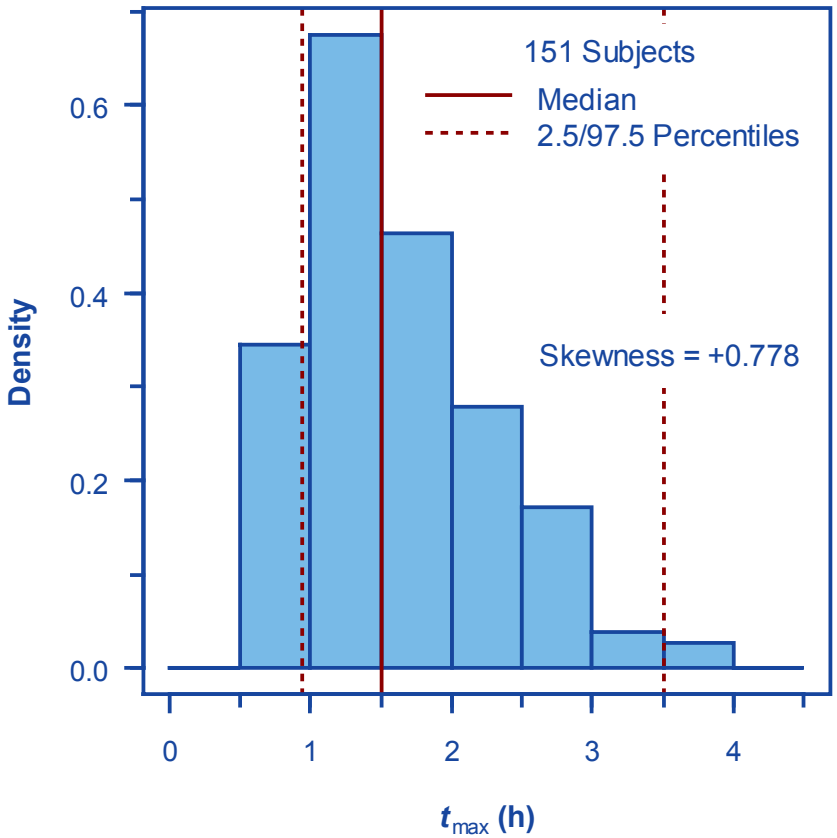
Calculating a ratio, *e.g.*, a percentage according to the EMA's product-specific guidances, is statistically flawed from the start



Same model as before; sampling every ten minutes ≤ 2 hours, 2.25, 2.5, 3, 3.5, 4, 6, 9, 12, and 16 hours

Statistical Properties of t_{max}

» The positive bias of T_{max} increase[s] together with the observational error. This result can be attributed to the asymmetry of the observed centrations around the peak. The concentrations rise more steeply before the peak than they decline following the true maximum response. Consequently, it is more likely that large observed concentrations occur after than before the true peak time. «



Tóthfálusi L, Endrényi L. *Estimation of C_{max} and T_{max} in Populations After Single and Multiple Drug Administration*. J Pharmacokin Pharmacodyn. 2003; 30(5): 363–85. doi:10.1023/b:jopa.0000008159.97748.09.

Simulations

2,500 studies, one-compartment model, three treatments: R (t_{max} 1.0 h), T_1 (t_{max} 0.8 h), T_2 (t_{max} 1.2 h), sampling every five minutes until two hours, 2.25, 2.5, 3, 3.5, 4, 6, 9, 12, 16 h

- The ≤ 20% difference criterion is <u>not a valid statistical test</u>
  - Hence, we cannot access the Type I Error
  - On the average we expect 50% of studies to pass the criterion
- If we pre-specify a clinically relevant difference of 0.2 h and apply the common confidence interval inclusion approach

$$\theta_1 = -\Delta \text{ and } \theta_2 = +\Delta$$

$$H_0 : \mu_T - \mu_R \not\subset \{\theta_1, \theta_2\} \text{ vs } H_1 : \theta_1 < \mu_T - \mu_R < \theta_2$$

by a nonparametric method, we could assess the Type I Error; since t_{max} (T_1) = t_{max} (R) − Δ and t_{max} (T_2) = t_{max} (R) + Δ we expect 5% of studies to pass

Simulations

Results

Treatment	Skewness	Range
Reference	+0.674	0.2500 – 4.000
Test 1	+0.778	0.1667 – 3.500
Test 2	+0.750	0.3333 – 6.000

- Positive skewness confirmed result of the real studies (+0.778)
- 57.9% of T_1 and 55.0% of T_2 passed the $\leq$ 20% difference criterion, which is larger than the 50% we expected
- If we follow the 'logic' of the product-specific guidances, Δ would be twelve minutes – is that clinically relevant?
- The Type I Error in the nonparametric method is controlled (5.32% of T_1 passed and 3.68% of T_2); not significant >5%
- Are the ranges 'apparently' different?

Issues

An extremely tight sampling schedule is required

- In the simulations we required 34 time points
- What if we have to deal with a painkiller (t_{max} 30 minutes)?
  - Is Δ of six minutes really clinically relevant?
  - Sampling every two minutes is a logistic nightmare

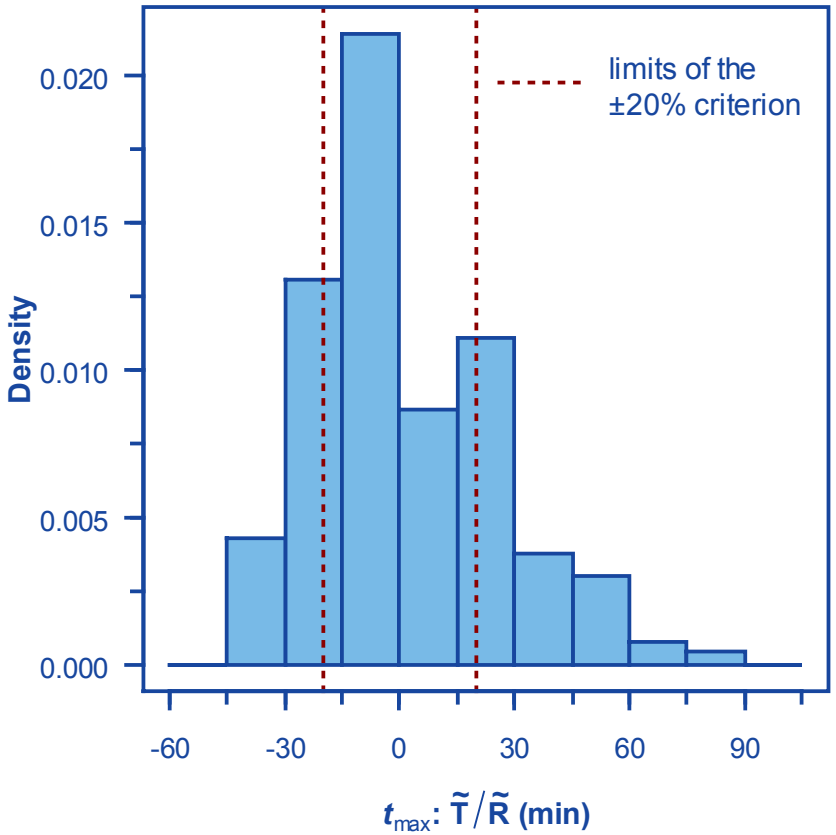
Sample size estimation is difficult

- Sufficient information about the drug (distribution, elimination) and the formulations (absorption) allowing to set up a suitable PK model
  - Not only the PK parameters themselves but also their variability would be required; a published population PK would come handy
  - Exploring different sampling schedules for various differences in t_{max}

Bootstrapping the Reference

600 mg IR ibuprofen, fasting state, 2×2×2 cross-over, 16 subjects (study* powered to ≥90% for C_{max}), sampling every 15 minutes until 2.5 hours; resampled t_{max} of the reference (!) in 10^5 simulations

- Empiric power 65.11%
- ≈60 subjects would be required to demonstrate BE of the reference to itself

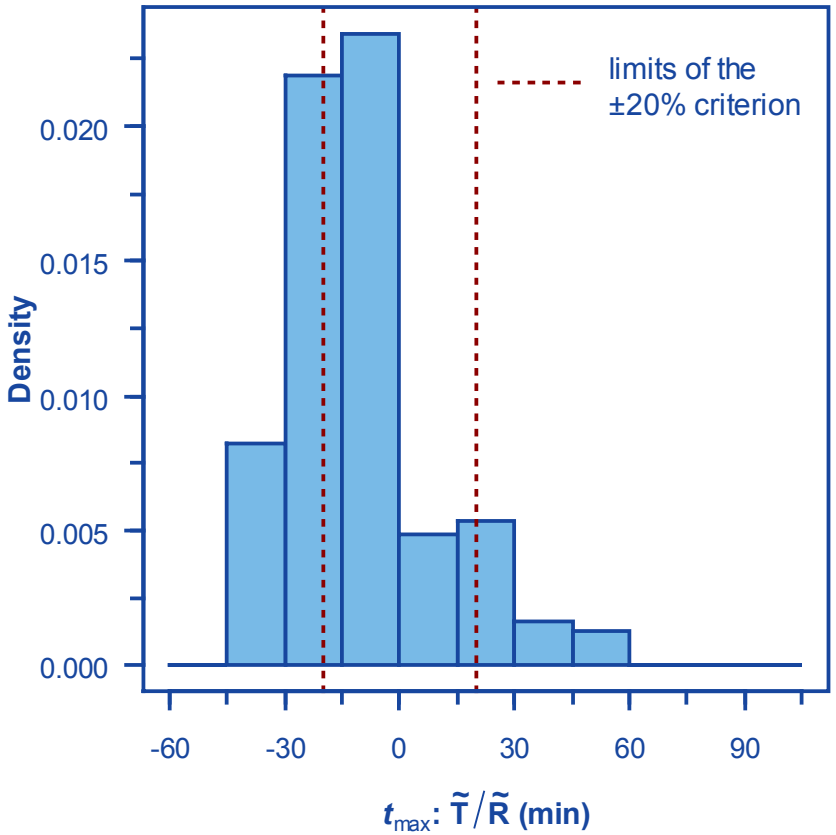


* Study performed in 1991. The generic product was approved in 1992 and is still on the market.

A slightly faster Test

Data of the reference but a test introduced which is eight minutes faster than the reference

- **Empiric power 51.60%**
- **That's hardly better than tossing a coin**
- **It would require ≈100 subjects to demonstrate BE**



Don't believe in Simulations?

400 mg IR ibuprofen, fasting state, 18 subjects, 2×2×2 cross-over, reference-replicated, washout three days*

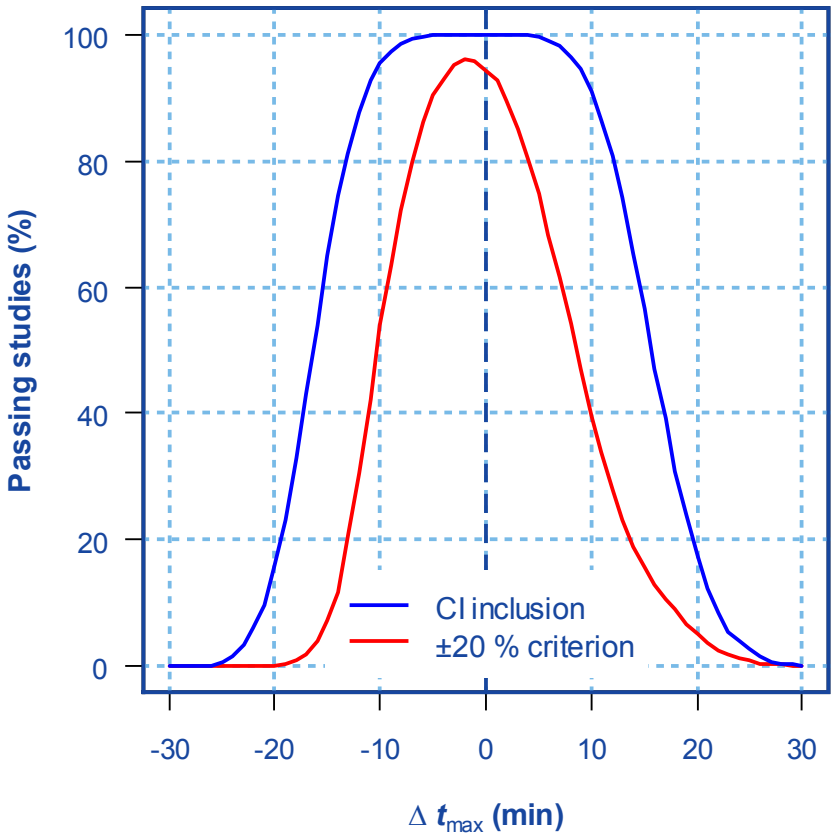
- Ranges of t_{max}
  - 1st administration: 0.25 – 4 hours
  - 2nd administration: 0.50 – 2 hours
- Insufficient sampling in the publication, therefore
  - Population PK model (one-compartment, no lag-time)
    - Reference based on the parameters of the PopPK model
    - Absorption rate constant of the Test increased to get a ten minutes earlier t_{max}
  - 'Sampling' every five minutes until 90 minutes, 1.75, 2, 2.25, 2.5, 3, 3.5, 4, 4.5, 5, 6, 8, and 12 hours
  - 2,500 studies simulated

* Wagener HH, Vögtle-Junkert U. *Intrasubject variability in bioequivalence studies illustrated by the example of ibuprofen.* Int J Clin Pharmacol Ther. 1996; 34(1): 21–31. PMID:8688993.

Don’t believe in Simulations?

Results

- 52.0% of simulated studies passed the ≤ 20% difference criterion
  - Asymmetrical power curve (shifted to the left); for any given power negative values are more likely to pass
  - Flawed due to calculating ratios with symmetrical limits
- 94.1% empiric power of the nonparametric CI inclusion approach with Δ 20 minutes
  - Almost symmetrical power curve



Summary

Paper does not blush

- Assessing t_{max} based on eyeballing 'apparent' differences of ranges is *bad science* and should be abandoned
  - There is no guarantee that by *looking* at reported ranges (what is 'apparent'?) an assessor will arrive at the same conclusions as the applicant – a great deal of discussions on its way
- The statement in the 2010 (IR) and 2014 (MR) guidelines

  » A [formal] statistical evaluation of t_{max} is not required «

  does not preclude to perform one
  - Only (!!) if clinically relevant, pre-specify an acceptance range for t_{max}; assess the 90% CI by an appropriate nonparametric method
- Calculating a ratio of data on an ordinal scale is simply *not allowed*
  - Thus, the ≤ 20% difference criterion in the EMA's recent product-specific guidances is flawed beyond repair

t_{max} Evaluation

Thank You!
Open Questions?

Helmut Schütz
BEBAC
1070 Vienna, Austria
helmut.schuetz@bebac.at
Institute of Medical Statistics, Medical University of Vienna
1090 Vienna, Austria
helmut.schuetz@meduniwien.ac.at