

# The Predictive Power of Dissolution and Alternatives to Full Bioequivalence

Helmut Schütz

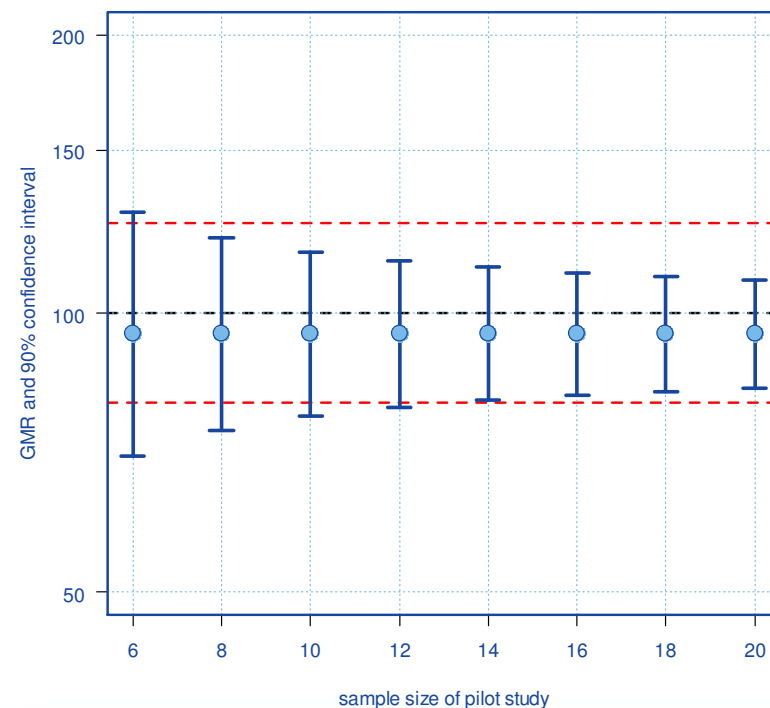


Wikimedia Commons • 2008 • Thomas Wolf • CCA-ShareAlike 3.0 Unported

# Pilot PK studies

## Power and size.

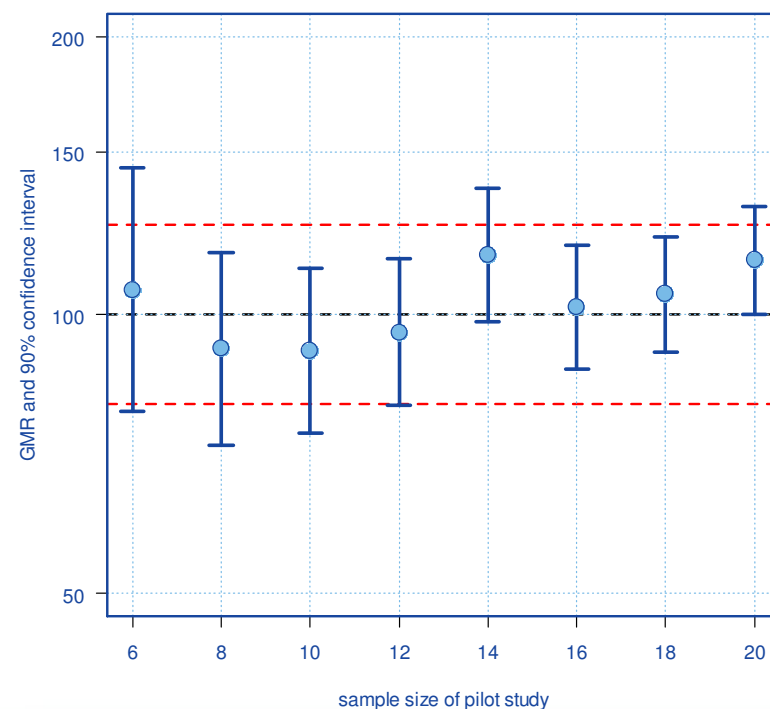
- By definition pilot studies are exploratory in nature; hence, they have no ‘power’.
  - It is tempting to keep the sample size as small as possible (six subjects were recommended in the last century).
  - Simulations with *GMR* 0.95.
  - The ‘true’ *GMR* lies with probability  $\alpha$  within the 90% CI.
  - With increasing sample size, the CI narrows.
  - It may be even possible to show BE (e.g., for the FDA).
  - Is it *that* simple?
    - Is it realistic to expect *identical GMRs* in *different* studies?



# Pilot PK studies

## Power and size.

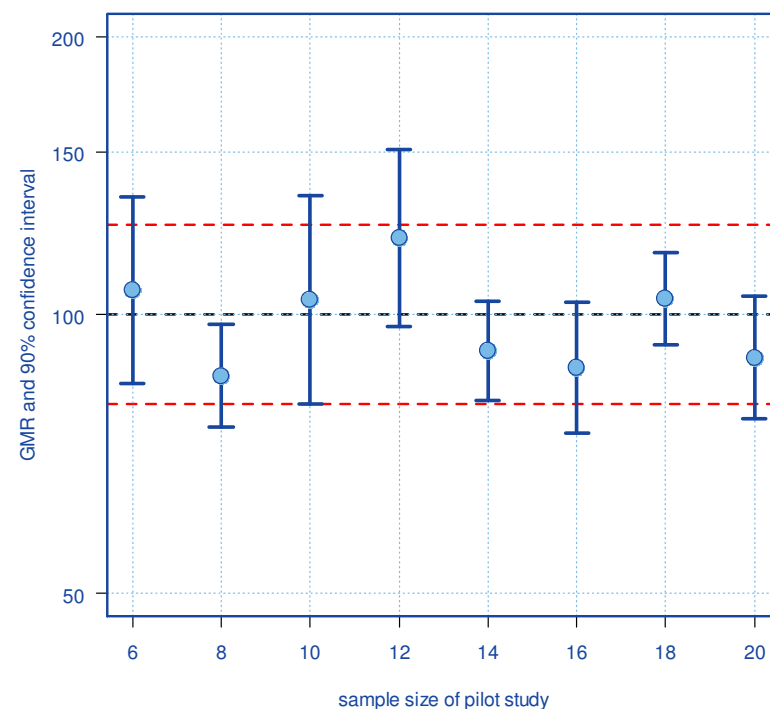
- Let us be more realistic:
  - Studies are done in different subjects. Therefore, we would expect that the *GMRs* will vary between studies as well.
  - If we compare different studies, the between-subject variability will hit (the *GMR* follows a log-normal distribution and I assumed that  $CV_{inter} = 2 \times CV_{intra}$ ).
  - Is this the end of the tunnel?
    - Could we expect that the within-subject *CV* will remain constant?
    - Maybe we should go even one step further.



# Pilot PK studies

## Power and size.

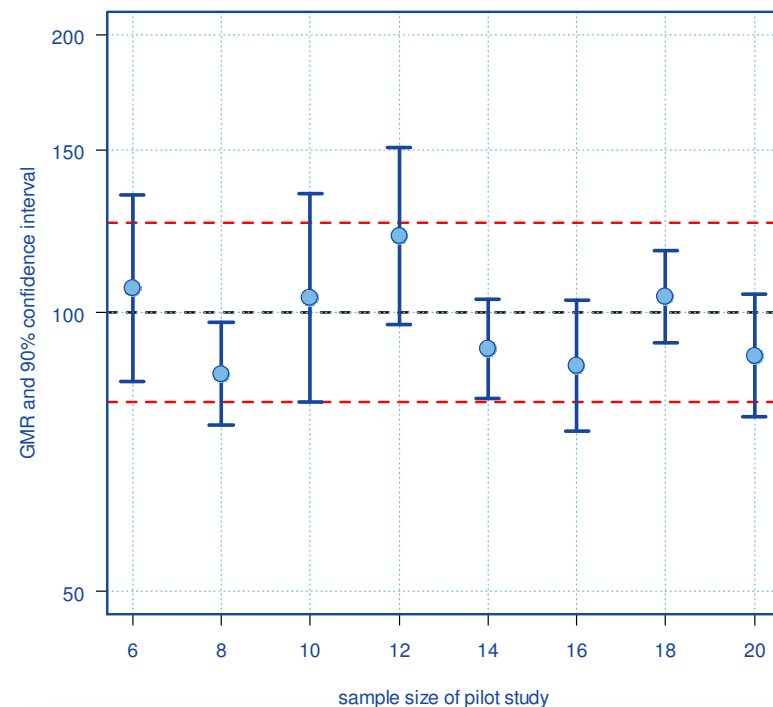
- Let us be more realistic:
  - Now we allow the within-subject variability also to vary (the variance follows a  $\chi^2$  distribution).
  - Unfortunately *this* is what we can expect from pilot studies.
    - Sometimes we get a false impression of low variability (e.g., with  $n = 8, 14, 18$ ).
    - Sometimes we get a false impression of almost perfectly matching products (e.g., with  $n = 10, 18$ ).
  - To get *reliable* estimates of both the *GMR* and the *CV* we would have to perform pilot studies which a *larger* than the pivotal one!



# Pilot PK studies

## Power and size.

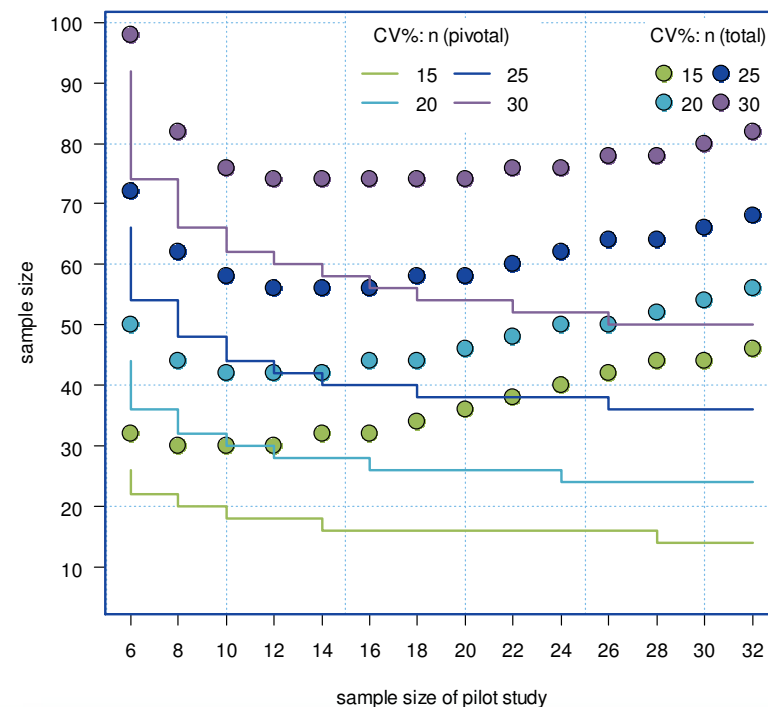
- Let us explore the details:
  - The *GMR* in the pilot study with 12 subjects is 121%.
    - Should we trash the formulation since the *GMR* is that close to the upper BE-limit – or not?
    - We know the true value: 95%!
  - The *CV* in the pilot study with 10 subjects is 32%.
    - Is this a HVD(P)?
    - We know the answer: No!
    - Particularly nasty. If we aim for reference-scaling (RSABE/ABEL) we would need a smaller sample size (compared to ABE). If in the pivotal the *CV* turns out to be <30% we would not be allowed to scale and will be *underpowered* for ABE.



# Pilot PK studies

## Sample size based on upper CL of the CV in the pilot

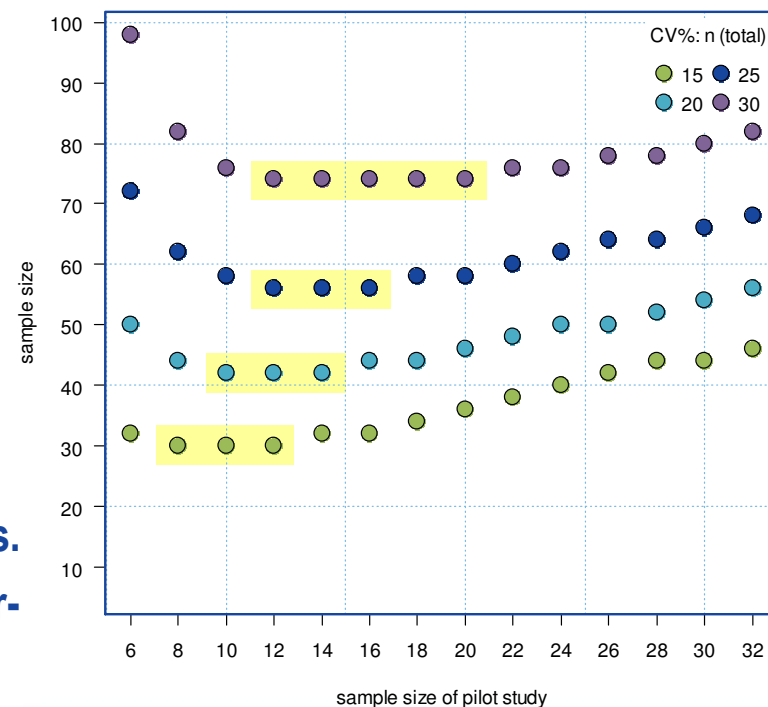
- The larger the pilot, the more precise the estimated CV.
  - Its upper CL will be more close to the estimate.
  - Hence, the sample size of the pivotal study will be *smaller*.
  - Very small pilots are practically use-  
less (due to the more imprecise CV).
  - Examples
    - CV 15%, pilot sample size 12.
      - » Upper CL of the CV 19.1%.
      - » Sample size of pivotal 18 (total 30).  
‘Carved in stone’: 12 (24).
    - CV 25%, pilot sample size 16.
      - » Upper CL of the CV 30.6%.
      - » Sample size of pivotal 40 (total 56).  
‘Carved in stone’: 28 (44).



# Pilot PK studies

## Sample size based on upper CL of the CV in the pilot

- The larger the pilot, the more precise the estimated CV.
  - However, we have to pay for both studies.
    - It seems that there are **minima** in the total sample sizes (dependent on the CV).
    - Can we conclude that there is an ‘ideal’ size of a pilot study? The smaller the better?
    - We also want to get a precise estimate of the *GMR*.
      - » Opt for a pilot size at the *upper end* of the ‘ideal’ range (or slightly larger).
    - If you suspect a HVD(P), perform the pilot in a replicate design to estimate the  $CV_{WR}$ .
      - » Some companies’ policy for HVDP(s) is a full replicate in at least 24 subjects.
    - Don’t use the ‘carved in stone’ CV and perform a power analysis (Session 4, part I).



# Pilot PK studies

## Does statistics help at all?

- To some extent, yes. See Fuglsang (2015) for further information.
- Caveats
  - The most critical value is the *GMR* – which is difficult to assess in a pilot study.
    - » Charles DiLiberti (2016) presented an example taking the worst case (pilot study in 12 subject; upper CL of CV 25% and *GMR* 0.95) into account. This would result in a ‘perfect’ pivotal sample size of 128 [*sic*] subjects.
  - Never assume perfectly matching products.
    - » The batch release spec’s are  $\pm 5\%$  of the declared content.
    - » Ask the QC lab about the accuracy and precision of the method (excellent ones have  $\sim 2.5\%$ ). The GL requires T- and R-batches deviating in content  $\leq 5\%$ . Even if you are extremely lucky to find batches with a *measured* content of 100% the *true* content may differ by  $100 - 100 (97.5 / 102.5) \sim 5\%$ .
  - Two-Stage Designs deal only with the CV. Allow for a safety margin of the *GMR*.

Fuglsang A. *Pilot and Repeat Trials as Development Tools Associated with Demonstration of Bioequivalence*.  
AAPS J. 2015; 17(3): 678–83. DOI 10.1208/s12248-015-9744-6.

DiLiberti C. *Adaptive Design Bioequivalence Studies: Controlling the Type 1 Error Rate While Preserving Power*.  
Rockville, 14–16 September, 2016: *The Global Bioequivalence Harmonization Initiative: EUFEPS/AAPS 2<sup>nd</sup> International Workshop*.



# The Predictive Power of Dissolution and Alternatives to Full Bioequivalence

**Thank You!**  
*Open Questions?*



**Helmut Schütz**  
**BEBAC**

Consultancy Services for  
Bioequivalence and Bioavailability Studies  
1070 Vienna, Austria  
[helmut.schuetz@bebac.at](mailto:helmut.schuetz@bebac.at)