How to design a pilot study

extrapolation of results

Helmut Schütz

Review of Guidelines

Minimum sample size (pivotal study)

- 12 WHO, EU, CAN, NZ, AUS, AR, MZ, ASEAN States, RSA, Russia ('Red Book'), EAEU, Ukraine
- 12 USA *'A pilot study that documents BE can be appropriate, provided its design and execution are suitable and a sufficient number of subjects (e.g., 12) have completed the study.'*
- 18 Russia (2008)
- 20 RSA (MR formulations)
- 24 Saudia Arabia (12 to 24 if statistically justifiable)
- 24 Brazil; USA (replicate designs intended for RSABE)
- 24 EU (RTR|TRT replicate designs intended for ABEL)
- 'Sufficient number' Japan
- 'Adequate' India

Review of Guidelines

Maximum sample size (pivotal study)

- Generally *not* specified (decided by IEC/IRB and/or local Authorities).
- ICH E9, Section 3.5 states:

  The number of subjects in a clinical trial should always be large enough to provide a reliable answer to the questions addressed.

Sample size (pilot study)

- Is ICH E9 also applicable?
- If yes (likely), what is a 'reliable' answer?

Power and Sample Size

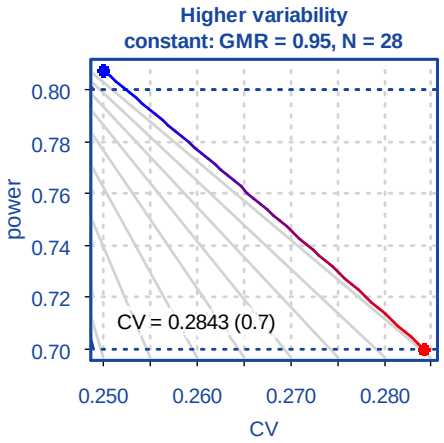
Which sample size is ‘large enough’?

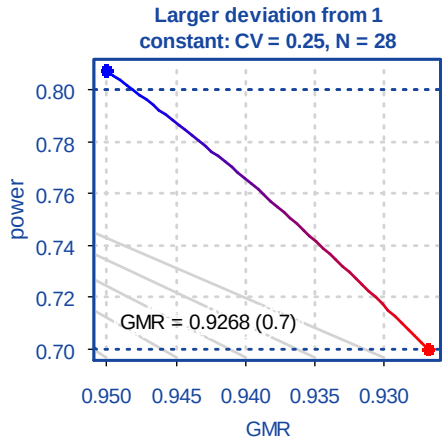
- Most guidelines recommend 80 – 90% power for pivotal studies.
  - EMA Appropriate sample size calculation [*sic*].
    Sample size depends on α (fixed), BE-limits (fixed),
    Δ (assumed), and desired power.
  - If a study is planned for ≤70% power, problems with the ethics committee are possible (ICH E9).
  - If a study is planned for >90% power (especially with low variability drugs), additional problems with regulators are possible (‘forced bioequivalence’).
  - Some subjects (‘alternates’) may be added to the estimated sample size according to the expected dropout-rate – especially for studies with more than two periods or multiple-dose studies.
- According to ICH E9 a sensitivity analysis is mandatory to explore the impact on power if values deviate from assumptions.

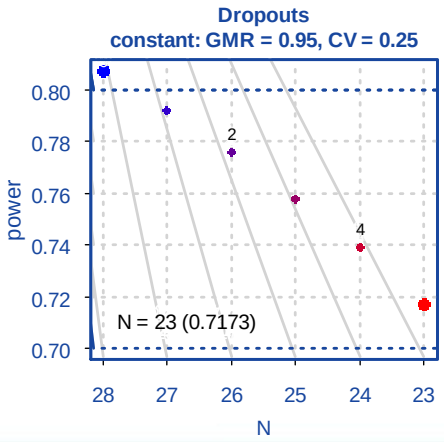
Power Analysis

Example 2×2×2, ABE

- Assumed *GMR* 0.95, CV_w 0.25, desired power 0.8, min. acceptable power 0.7.
  - Sample size 28 (power 0.807)
  - CV_w can increase to 0.284 (rel. +14%)
  - *GMR* can decrease to 0.927 (rel. –2.4%)
  - 5 drop-outs acceptable (rel. –18%)
  - Most critical is the *GMR*!



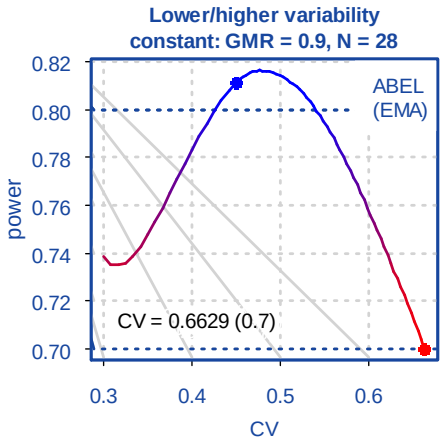


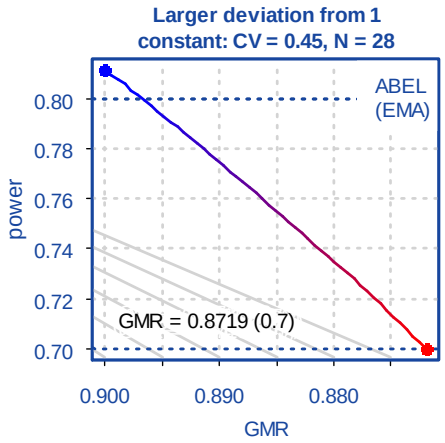


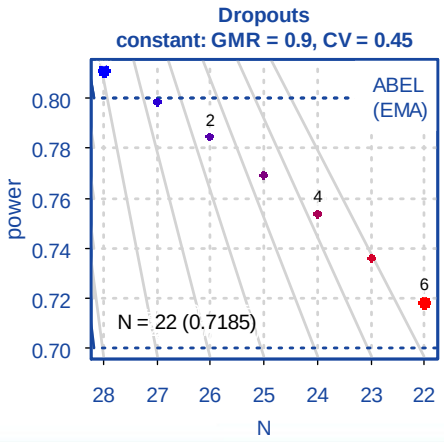
Power Analysis

Example 2×2×4, ABEL

- Assumed *GMR* 0.90, CV_{wR} 0.45, desired power 0.8, min. acceptable power 0.7.
  - Sample size 28 (power 0.811)
  - CV_w can increase to 0.663 (rel. +47%)
  - *GMR* can decrease to 0.872 (rel. –3.1%)
  - 6 drop-outs acceptable (rel. –21%)
  - Most critical is the *GMR*!



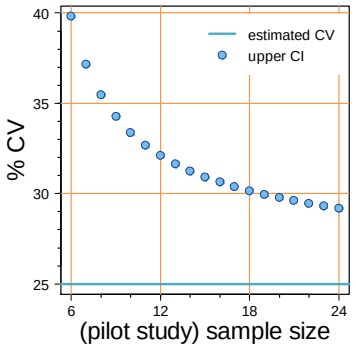




Dealing with Uncertainty

Nothing is ‘carved in stone’

- Never assume perfectly matching products.
  - Generally a Δ of *not* better than 5% should be assumed (*GMR* 0.9500 – 1.0526).
  - For HVD(P)s do not assume a Δ of <10% (*GMR* 0.9000 – 1.1111).
- Do not use the *CV* but one of its confidence limits.
  - Suggested α 0.2 (here: the producer’s risk).
  - For ABE the upper CL.
  - For reference-scaling the lower or upper CL.
- Precision of estimates.
  - Improves with n^2.
  - In order to double the precision one has to quadruple the sample size.



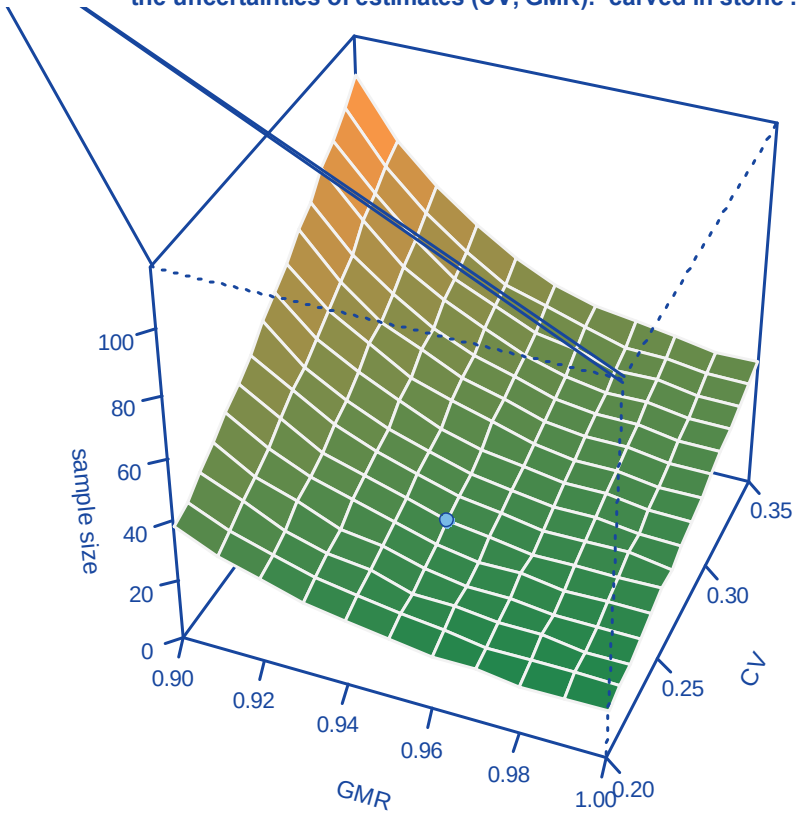
Dealing with Uncertainty

Precision

- 2×2×2 pilot studies (different sample sizes), *GMR* 0.95, CV_w 0.25
  - 80% confidence intervals (20% producer's risk) of *GMR* and CV_w.
    - *n* 12: *GMR* 0.8276 – 1.0905 (90% CI: 75.94 – 118.85%)
      CV_w 0.1835 – 0.4078
    - *n* 16: *GMR* 0.8450 – 1.0680 (90% CI: 78.82 – 114.50%)
      CV_w 0.1910 – 0.3713
    - *n* 24: *GMR* 0.8648 – 1.0435 (90% CI: 81.98 – 110.09%)
      CV_w 0.2002 – 0.3379
  - Sample sizes of the pivotal study (80% power) based on worst case scenarios (lower CL of *GMR* and upper CL of CV_w).
    - *n* 12: 1,656
    - *n* 16: 536
    - *n* 24: 222
  - Alternative: Bayesian method.

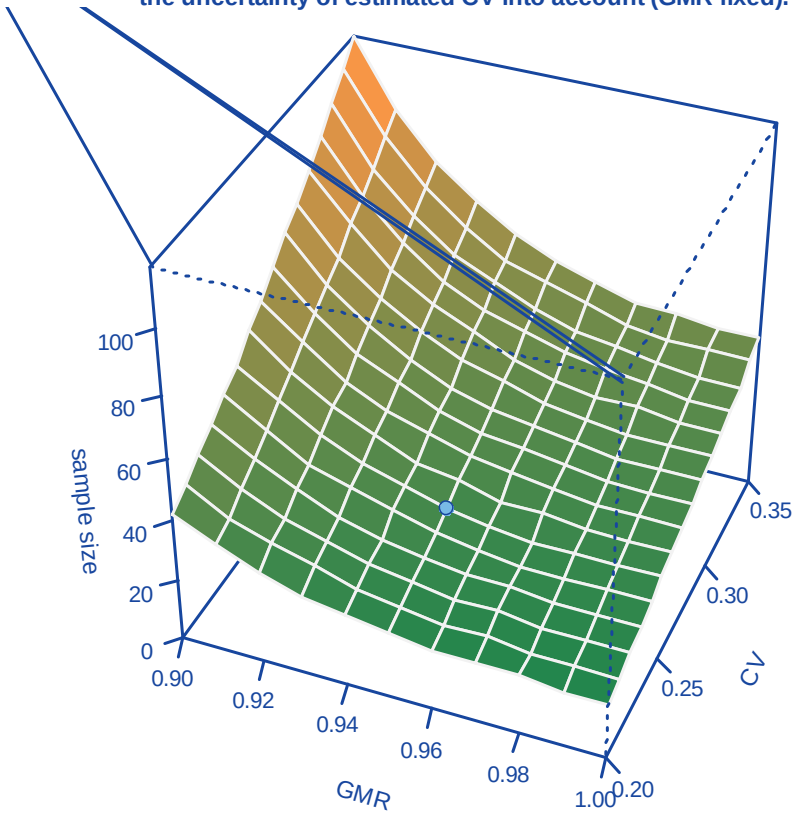
Bayesian Method

Pivotal study (80% power) designed on results of a 2x2x2 pilot study with 16 subjects ignoring the uncertainties of estimates (CV, GMR): 'carved in stone'.



sample size for GMR 0.95 and CV 0.25: 28

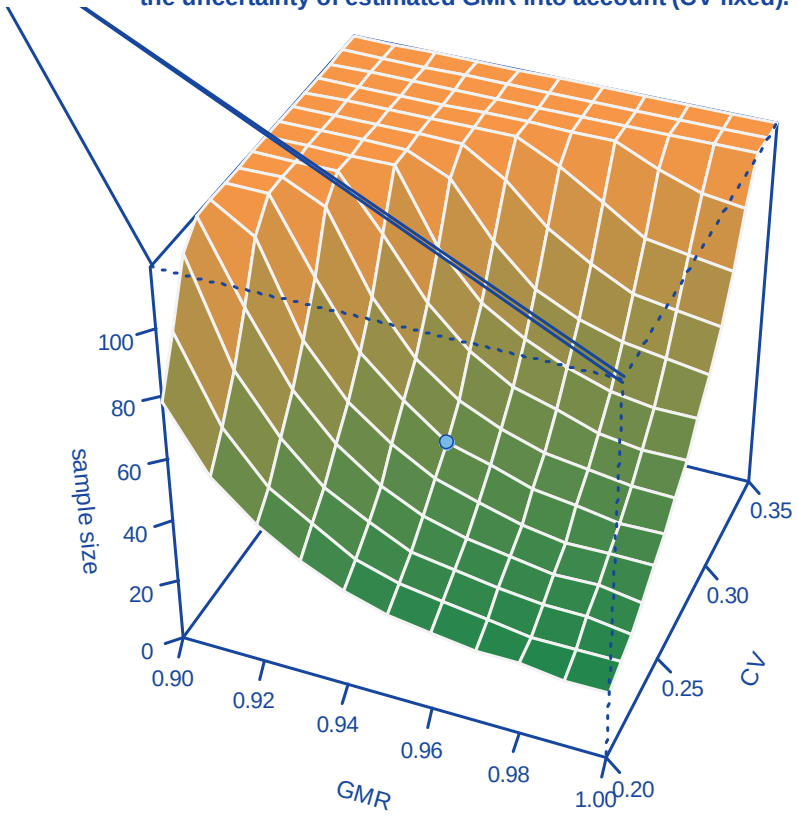
Pivotal study (80% power) designed on results of a 2x2x2 pilot study with 16 subjects taking the uncertainty of estimated CV into account (GMR fixed).



sample size for GMR 0.95 and CV 0.25: 32

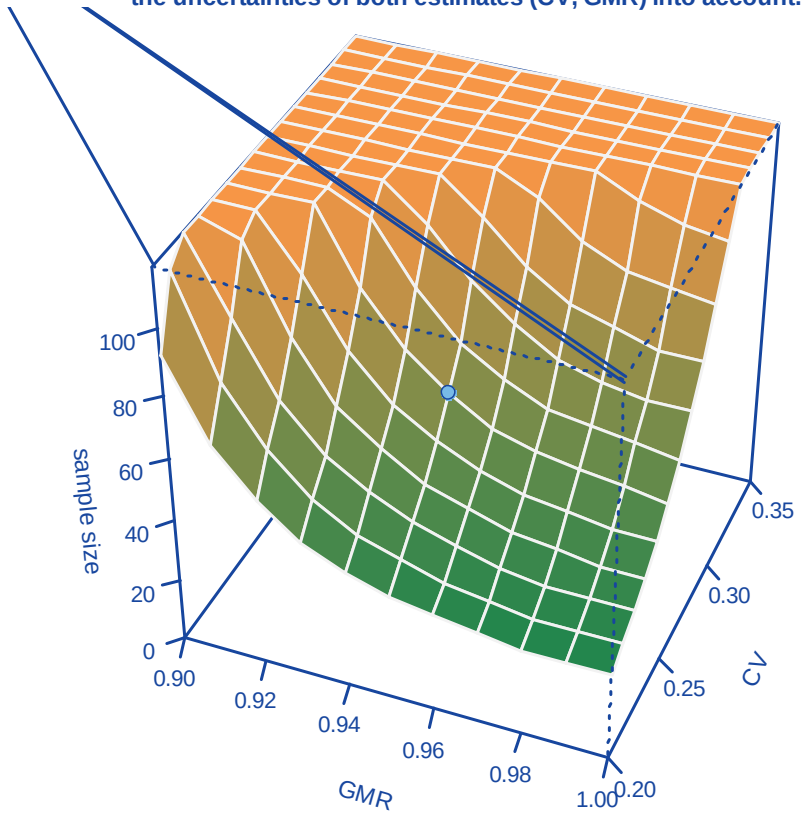
Bayesian Method

Pivotal study (80% power) designed on results of a 2x2x2 pilot study with 16 subjects taking the uncertainty of estimated GMR into account (CV fixed).



sample size for GMR 0.95 and CV 0.25: 54

Pivotal study (80% power) designed on results of a 2x2x2 pilot study with 16 subjects taking the uncertainties of both estimates (CV, GMR) into account.



sample size for GMR 0.95 and CV 0.25: 70

Problems

The EMA's 'appropriate sample size calculation'

- The purpose of a pilot study (amongst others) is to obtain estimates of the *GMR* and *CV* which can be used to design the pivotal study.
- In a strict sense it is not possible to demonstrate bioequivalence in a pilot study which is – by definition – exploratory.
- However, in the past some agencies (Scandinavian countries, Germany) accepted pilot studies as evidence of BE if stated as such in the protocol.
  - Repeating a passing pilot (even in a larger sample size) may fail by pure chance (producer's risk = 1 – power).
  - Hence, this approach was considered unethical.
- Nowadays, European regulatory agencies seemingly are more strict (follow the 'cook book').

Still acceptable for the FDA…

Remedies, Outlook

Pilot study

- For applicants
  - Sample size as large as the budget allows.
    - Increases the precision of estimates.
    - Adjusting for the uncertainty of the *GMR* (even with the Bayesian method) leads to sample sizes of the pivotal study which likely are not feasible.
    - Take all available information about the *GMR* into account (*e.g.*, from *IVIVC*) but always allow for a safety margin (don't be overly optimistic).
  - For ABE consider a Two-Stage Design.
    - Adjusts the sample size based on the *CV* observed in the first stage.
    - Do not add more subjects in the second stage ('in order to compensate for potential loss in power due to dropouts). Use the re-estimated sample size; otherwise the Type I Error may be inflated.
    - Adjusting for the observed *GMR* is generally not possible (compromises power).
    - Include a futility criterion for early stopping.

Remedies, Outlook

Pilot study

- For applicants
  - Reference-scaling (ABEL)
    - If the expected CV_{wR} is within 30 – 50% and the actual CV_{wR} is larger, power increases (more expansion of limits).
    - Some companies have a policy for pilot studies:
      Full replicate, 36 subjects.
    - Even if the pivotal study is planned as a partial replicate design (RRT|RTR|TRR), perform the pilot in a full replicate in order to additionally estimate CV_{wT}.
      If $CV_{wT} < CV_{wR}$ there will be incentive in the sample size.
      Example
      - » CV_{wT} 35%, CV_{wR} 50% observed in the full replicate pilot.
        Sample size for a partial replicate design 33.
      - » If the pilot was performed in a partial replicate (no information about CV_{wT}) one has to *assume* that $CV_{wT} = CV_{wR}$.
        Sample size for a partial replicate design 39.

Remedies, Outlook

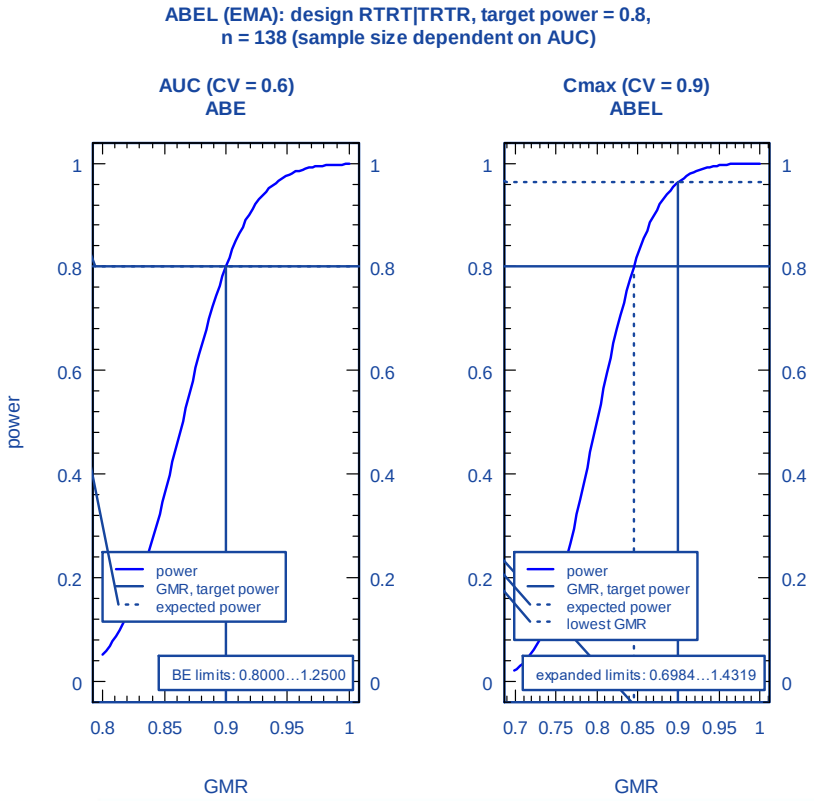
Pilot study

- For applicants
  - Demonstrating bioequivalence in the pilot.
    - State the intention unambigously in the protocol.
    - Give a justification and concentrate on ethics rather than economics.
    - Consider a scientific advice in a ‘difficult’ member state (*e.g.*, Spain, The Netherlands, France).

Remedies, Outlook

Pivotal study

- For applicants
  - The EMA's approach of allowing reference-scaling only for C_{max} has the side effect of accepting products which large deviations if *AUC* is highly variable as well.
    - The sample size depends on the variability of *AUC* which has to be assessed by ABE. Example:
      - » Target power 80%, *GMR* 0.9 (both PK metrics), $CV_{wT} = CV_{wR}$ 0.6 (*AUC*), 0.9 (C_{max}).
      - » With 138 subjects required for *AUC*, products with a *GMR* of 0.846 of C_{max} will pass ABEL.



Remedies, Outlook

Pilot study

- For regulatory agencies
  - Reconsider accepting BE demonstrated in a pilot study.
    - Example
      - » Pilot: *n* 24, *GMR* 0.95, CV_w 0.25, 90% CI 81.98 – 110.09%
      - » Pivotal: *n* 28, power 80.7% (*i.e.*, risk of failure 19.3%)
  - Elastic clause in the BE GL (4.1.8 Evaluation – Presentation of data)

    If […] multiple studies have been performed some of which demonstrate BE and some of which do not, the body of evidence must be considered as a whole. Only relevant studies, as defined in section 4.1, need be considered. The existence of a study which demonstrates BE does not mean that those which do not can be ignored. The applicant should thoroughly discuss the results and justify the claim that BE has been demonstrated. Alternatively, when relevant, a combined analysis of all studies can be provided in addition to the individual study analyses. It is not acceptable to pool together studies which fail to demonstrate BE in the absence of a study that does.

Remedies, Outlook

Pivotal study

- For regulatory agencies
  - Reconsider accepting reference-scaling also for *AUC*.
    - Was discussed in the Concept Paper 2006 (removed from the EMA's website; available at: http://bebac.at/downloads/14723106en.pdf) and the 2nd International Conference of the Global Bioequivalence Harmonization Initiative (Rockville, September 2016).
    - RSABE acceptable for the FDA.
    - ABEL acceptable for Health Canada (expanded limits up to 66.67 – 150.00%).
    - In June 2017 the WHO opened in pilot phase allowing scaling for *AUC* on a case-by-case basis. 4-period full replicate design mandatory 'in order to assess the variability associated with each product'.
    - Current practice leads to approval of products with large Δ in C_{max}. Although technically valid, is this really desirable?

How to design a pilot study

Thank You!
Open Questions?

Helmut Schütz
BEBAC
Consultancy Services for
Bioequivalence and Bioavailability Studies
1070 Vienna, Austria
helmut.schuetz@bebac.at