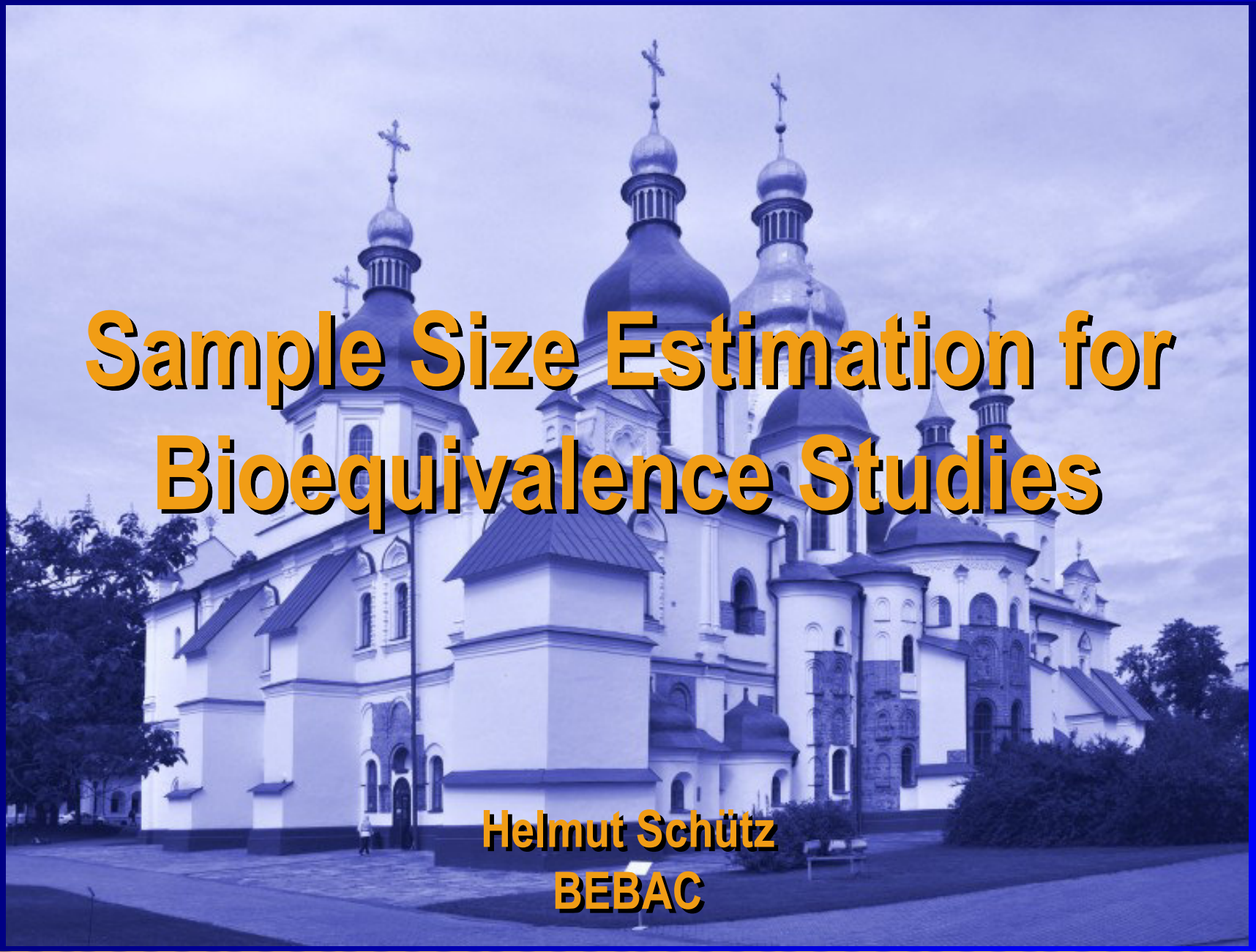


Sample Size Estimation for Bioequivalence Studies

Helmut Schütz
BEBAC

α and β

- All formal decisions are subjected to two 'types' of error:
  - α: Probability of Type I Error (aka Risk Type I)
  - β: Probability of Type II Error (aka Risk Type II)

  Example from the justice system – which presumes that the defendant is *not* guilty:

Verdict	Defendant innocent	Defendant guilty
Presumption of innocence rejected guilty	Error type I	Correct
Presumption of innocence accepted not guilty	Correct	Error type II

α and β

- In statistical terminology

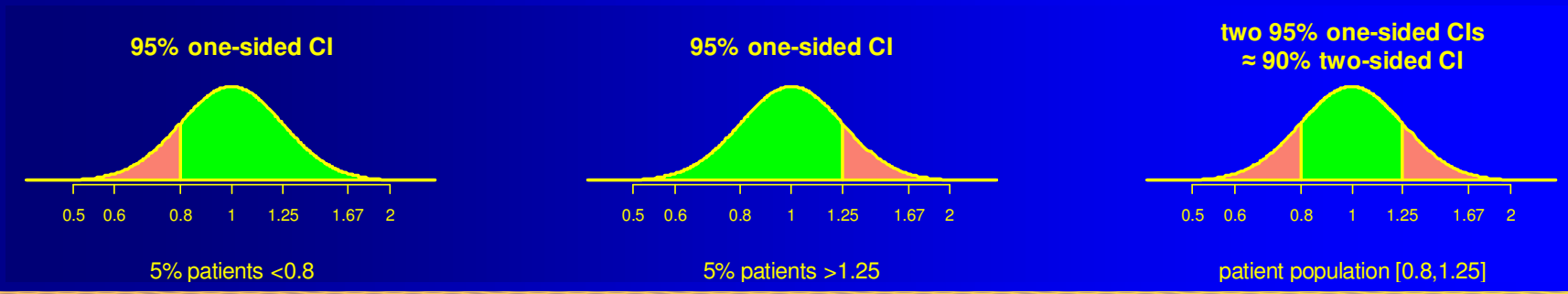
Decision	Null hypothesis true	Null hypothesis false
Null hypothesis rejected	Error type I	Correct (H_a)
Failed to reject null hypothesis	Correct (H_0)	Error type II

- In BE-testing the Null hypothesis is bioinequivalence ($\mu_1 \neq \mu_2$)!

Decision	Null hypothesis true	Null hypothesis false
Null hypothesis rejected	Patient's risk	Correct (BE)
Failed to reject null hypothesis	Correct (not BE)	Producer's risk

α …

- Patient's Risk to be treated with an inequivalent formulation (H_0 falsely rejected)
  - BA of the test compared to reference in a *particular* patient is risky ***either*** below 80% ***or*** above 125%.
  - If we keep the risk of *particular patients* at α 0.05 (5%), the risk of the entire *population of patients* (<80% ***and*** >125%) is 2α (10%) – expressed as a confidence interval: 100(1 – 2α) = 90%.



… and β

- **Producer's Risk to get no approval of an equivalent formulation** (H_0 falsely not rejected)
  - *Fixed* in study planning to ≤0.2 (20%), where power = 1 − β = ≥80%
  - If power is set to 80%,
    one out of five studies will fail just by chance!

α 0.05	BE
not BE	β 0.20

0.20 = 1/5

  - *A posteriori* (*post hoc*) power does not make sense!
    Either a study has demonstrated BE or not.

Sample Size (Guidelines)

- **Minimum**
  - 12 WHO, EU, CAN, NZ, AUS, AR, MZ, ASEAN States, RSA, Russia ("Red Book"), 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 (in replicate designs intended for RSABE)
  - 'Sufficient number' Japan

Sample Size (Guidelines)

- Maximum
  - 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.'

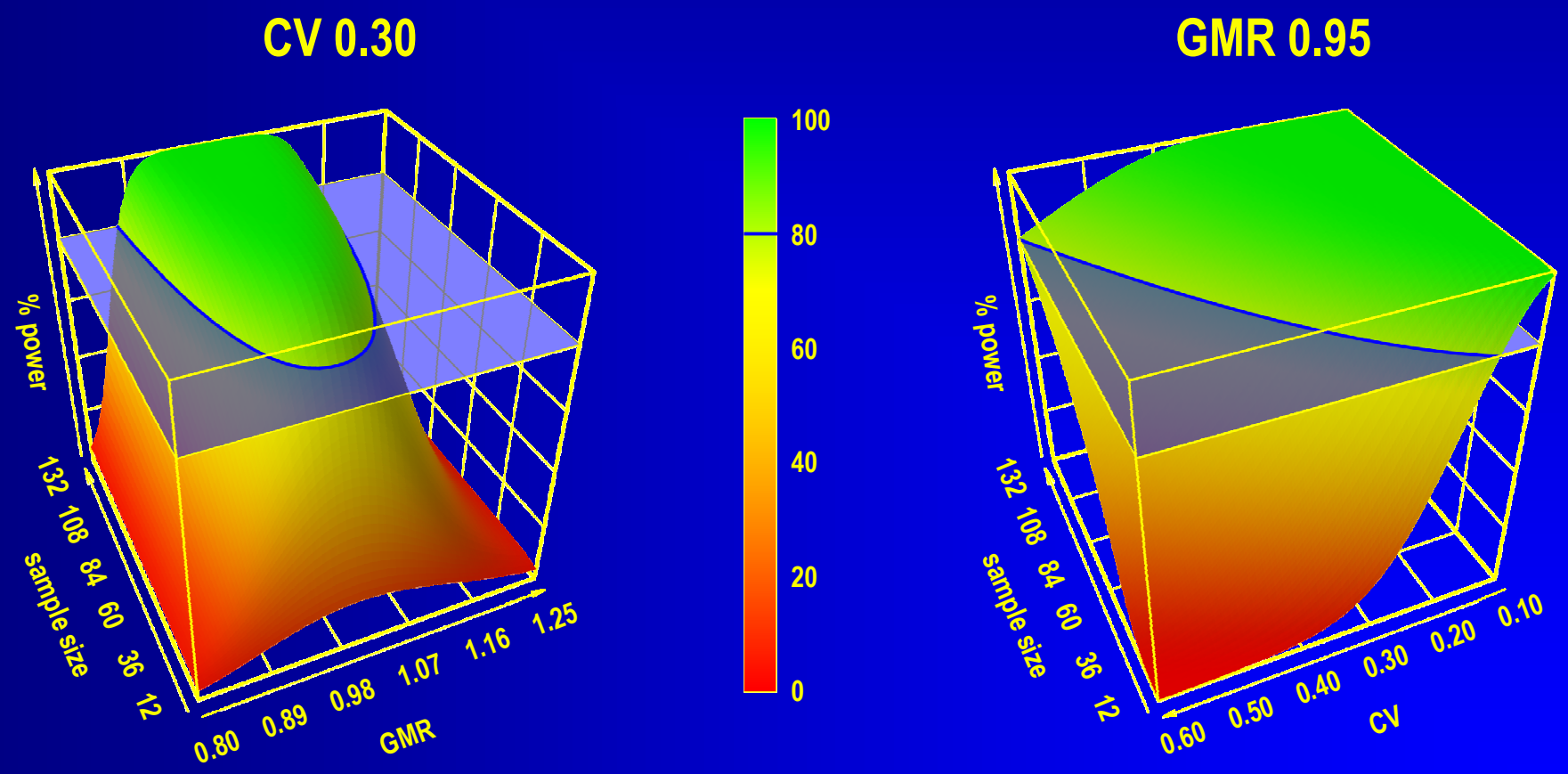
No Sample Size 'Calculation'

- It is not possible to *directly* calculate the required sample size.
- Power is calculated instead;
  the smallest sample size which fulfills the minimum target power is used.
  - Example:
    α 0.05, target power 80% (β 0.2),
    expected T/R 0.95, CV_{intra} 30% →
    minimum sample size 39 (power 80.6%),
    rounded *up* to the next even number in a 2×2×2 study (power 81.6%).

n	Power (%)
36	77.24
37	78.39
38	79.53
39	80.56
40	81.58

Power Surfaces

2×2×2 crossover study, target power 80%



Power *vs.* Sample Size

- How many subjects are enough?
  - Most guidelines recommend 80 – 90% 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') should be added to the estimated sample size according to the expected drop-out rate – especially for studies with more than two periods or multiple-dose studies.

Hierarchy of Designs

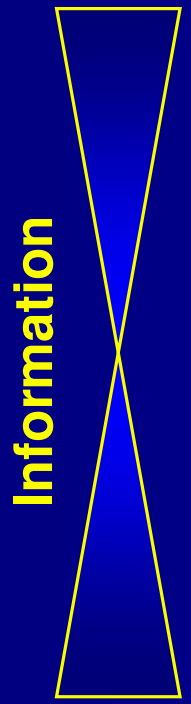
- The more 'sophisticated' a design is, the more information can be extracted.
  - Hierarchy of designs:

    Full replicate (TRTR | RTRT or TRT | RTR),

    Partial replicate (TRR | RTR | RRT)

    Standard 2×2×2 cross-over (TR | RT)

    Parallel (R | T)



  - Variances which can be estimated:

    Parallel: total variance (between + within subjects)

    2×2×2 Xover: + between, within subjects

    Partial replicate: + within subjects of reference

    Full replicate: + within subjects of reference and test

Coefficient(s) of Variation

- From any design one gets variances of designs which are lower in the hierarchy as well.
  - Total CV% from a 2×2×2 crossover used in planning a parallel design study:
    - Intra-subject CV% (within)
    - Inter-subject CV% (between)
    - Total CV% (pooled)

$$CV_{intra}\,\% = 100 \cdot \sqrt{e^{MSE_W} - 1}$$

$$CV_{inter}\,\% = 100 \cdot \sqrt{e^{\frac{MSE_B - MSE_W}{2}} - 1}$$

$$CV_{total}\,\% = 100 \cdot \sqrt{e^{\frac{MSE_B + MSE_W}{2}} - 1}$$

Coefficient(s) of Variation

- However, CVs of *higher* design levels not available.
  - If only mean ± SD of reference is available…
    - 'Rule of thumb' $CV_{intra} \approx 60\%$ of CV_{total} not correct.
    - A crossover study must not be planned based on CV_{total}.
    - Examples (2×2×2 studies)

drug	formulation	design	n	PK metric	CV_{intra}	CV_{inter}	CV_{total}
methylphenidate	biphasic MR	SD	12	AUC_{0-t}	7.0	19.1	20.4
laroxetine	XR	MD	32	$AUC_{0-\tau}$	25.2	55.1	62.1
lansoprazole	DR	SD	47	C_{max}	47.0	25.1	54.6

    - Pilot study unavoidable, unless a TSD is used.

Tools

- Sample Size Tables (Phillips, Diletti, Hauschke, Chow, Julious, …)
- Approximations (Diletti, Chow, Julious, …)
- General purpose (SAS, S+, *R*, StaTable, …) and specialized software (nQuery Advisor, PASS, FARTSSIE, StudySize, …)
- Exact method (*R*-package *PowerTOST*)
  Currently only *PowerTOST* provides sample size estimation for reference-scaling (EMA, FDA) and FDA's method for NTIDs.

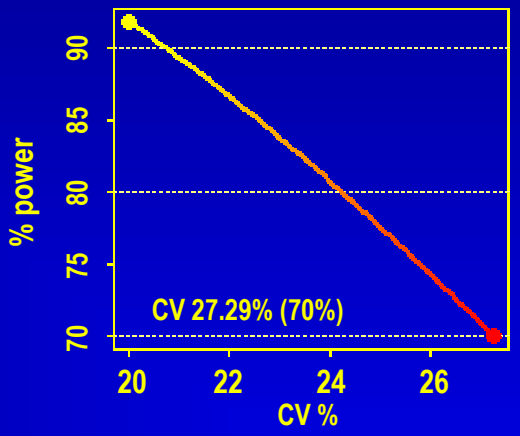
Sensitivity Analysis

- ICH E9 (1998)
  - Section 3.5 Sample Size, paragraph 3
    - The method by which the sample size is calculated should be given in the protocol […]. The basis of these estimates should also be given.
    - It is important to investigate the sensitivity of the sample size estimate to a variety of deviations from these assumptions and this may be facilitated by providing a range of sample sizes appropriate for a reasonable range of deviations from assumptions.
    - In confirmatory trials, assumptions should normally be based on published data or on the results of earlier trials.

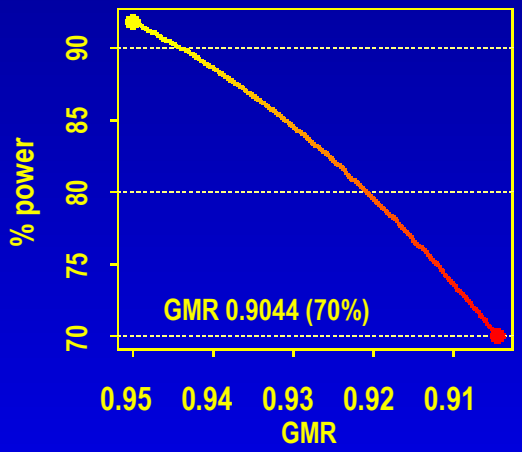
Sensitivity Analysis

- Example
  - CV 20%, GMR 0.95, AR 80 – 125%, target power 90%, min. acceptable 70%.
  - Estimated sample size: 26 (91.76% power)
  - Acceptable relative deviations:
    - CV +36.4%
    - GMR –4.80%
    - n –38.5%

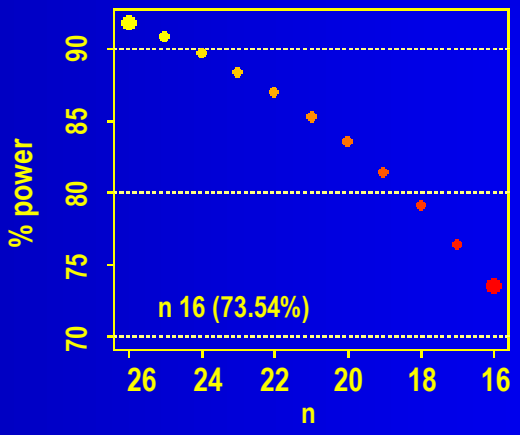
Higher variability
GMR fixed at 0.95



Higher deviation of GMR
CV fixed at 20%



Influence of dropouts
GMR 0.95, CV 20%



Sensitivity Analysis

- Has to be done *before* the study (*a priori*).
- The Myth of retrospective (*a posteriori*) Power:[1,2,3]
  - High power *does not* further support the claim of already demonstrated bioequivalence.
  - Low power *does not* invalidate the conclusion of claimed bioequivalence.

1. Lenth RV (2000)
   Two Sample-Size Practices that I don't recommend
   http://www.math.uiowa.edu/~rlenth/Power/2badHabits.pdf
2. Hoenig JM, Heisey DM (2001)
   The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis
   The American Statistician 55(1):19–24
   http://www.vims.edu/people/hoenig_jm/pubs/hoenig2.pdf
3. Bacchetti P (2010)
   Current sample size conventions: Flaws, harms, and alternatives
   BMC Medicine 8(17) DOI: 10.1186/1741-7015-8-17

Thank You!

Sample Size Estimation for BE Studies

Questions?

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

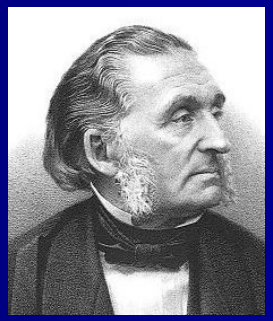
To bear in Remembrance...

Power. That which statisticians are always calculating but never have.

Power: That which is wielded by the priesthood of clinical trials, the statisticians, and a stick which they use to beta their colleagues.

Power Calculation – A guess masquerading as mathematics.

Stephen Senn



You should treat as many patients as possible with the new drugs while they still have the power to heal.

Armand Trousseau