Analytical method transfer: proposals for the location-scale approach and tolerance intervals

Cornelia Frömke, Ludwig A. Hothorn, Michael Schneider

Institute of Biometry, Hannover Medical School
Institute of Biostatistics, Leibniz University Hannover

27.-29.09.2010

The method transfer

Procedure

- an analytical method is established in laboratory a (*sender*)
- method shall be transferred to another laboratory b (*receiver*)
- Aim: to prove similarity of both laboratories. Hence:
- samples $i = 1, \ldots, n$ for analysis are created; each sample is split in two subsamples
- subsamples are analyzed in lab j, where $j = a, b$
- observations are denoted as x_{ij}

The method transfer

Data

- paired data
- method comparison: different data scales possible
  method transfer: same data scales
- range of method should cover all future incoming samples
- here: no repeated measurements (sample with a specific concentration not repeated)

In general

- correlation > 50%
- location test is not robust to a systematic proportional difference - however, this is not expected in an analytical method transfer

Example

Lipase activity of 51 different product batches was measured in two laboratories over a time period of approximately 2 years (collaborative trial).

total data consists of $2n$ ($n = 51$)
paired observations

lab a (Ph.Eur.-u/g)	**lab b (Ph.Eur.-u/g)**
59017	55943
138940	120628
31992	31496
61741	62585
65451	67979
80447	67923
108324	95589
58359	54479
23250	21903
30627	30454
...	...

Statistical guidelines for method transfer

- FDA Guidance for Industry: Protocols for the Conduct of Method Transfer Study for Type C Medicated Feed Assay Methods
- EMEA Validation of Bioanalytical Method
- FDA Guidance for Industry: Bioanalytical Method Validation

Trueness and precision should be lower than 15%.

→ trueness and precision can be interpreteded as location and scale
→ margin of 15% is multiplicative - therefore we suggest ratios

Selection of current statistical approaches

Typically, similarity of two laboratories is transferred to equivalence of the laboratories

- *test on equivalence for trueness and precision*
  location and scale usually combined with IUT (Choudhary et al., 2005)
  or one single test for location and scale (Bradley et al., 1989)
- *tolerance/predication intervals*
  test for limits of tolerance intervals (Zhong et al., 2008)
  total error approaches (Hoffman et al., 2007, Rozet et al., 2009)
  prediction intervals (Carstensen et al., 2010)

→ single test for location and scale: interpretation is difficult (especially when transfer has failed)
→ of interest: separate confidence intervals for location and scale - separate and post-hoc interpretation possible

Two-sided $100(1-\alpha)\%$ tolerance interval (TI)

… which contains at least a proportion p of the population

Parametric TI for differences (Hahn and Meeker, 1991)

$TI_p^{diff} = \bar{d} \pm k_{1-\alpha;p,n} s_d,$

where $k_{1-\alpha;p,n}$ is tabulated in e.g. Hahn and Meeker (1991)

Nonparametric TI for differences (Hahn and Meeker, 1991)

$TI_{np}^{diff} = [d_{lower}; d_{upper}]$

Sort d_i in ascending order: $d_{(1)} \leq d_{(i)} \leq d_{(n)}$.
Cut off smallest and largest values, where number of values to be removed are tabulated in Hahn and Meeker (1991)

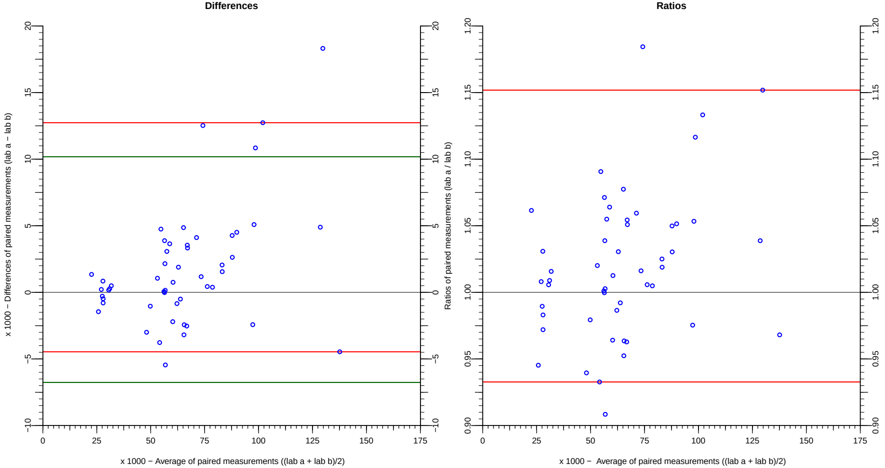
Modification: nonparametric TI for ratios

$TI_{np}^{ratio} = [r_{lower}; r_{upper}]$

Sort r_i in ascending order: $r_{(1)} \leq r_{(i)} \leq r_{(n)}$.
Cut off smallest and largest values, where number of values to be removed are tabulated in Hahn and Meeker (1991).

Lipase data: Bland-Altman plot

Bland-Altman plot (Bland et al., 1986) extended with 90% tolerance interval to contain 90% of the population



Location-scale approach

Basic idea

Similarity of laboratories = equivalence of location AND non-inferiority of scale by means of confidence intervals.
All intervals are marginal due to intersection union principle.

Hypotheses

Based on differences:
$H_0 : \{|\mu_a - \mu_b| \geq \delta\} \cup \{\sigma_a - \sigma_b \leq \varepsilon\}$ vs. $H_1 : \{|\mu_a - \mu_b| < \delta\} \cap \{\sigma_a - \sigma_b > \varepsilon\}$

Based on ratios:
$H_0 : \frac{\mu_a}{\mu_b} \leq \theta^{-1} \cup \frac{\mu_a}{\mu_b} \geq \theta \cup \frac{\sigma_a}{\sigma_b} \leq \lambda$ vs. $H_1 : \theta^{-1} < \frac{\mu_a}{\mu_b} < \theta \cap \frac{\sigma_a}{\sigma_b} > \lambda$

Note

Similarity thresholds: $\delta > 0, \varepsilon < 0, \theta > 1, \lambda < 1$.

Note: location: (1-2α)-confidence interval,
for scale: (1-α) lower confidence bound

Parametric CI's for difference

$(1-2\alpha)$ CI for the difference of means (Altman, 1990)

$$CI_{\bar{d}} = \bar{d} \pm t_{n-1,1-2\alpha}\sqrt{\frac{1}{n}}s_d,$$

where $t_{n-1,1-2\alpha}$ denotes the $(1-2\alpha)$-quantile of the t-distribution with $n-1$ d.o.f.

$(1-\alpha)$ lower confidence bound for difference of variances

$$CI_{s^2} = \left[s_a^2 - s_b^2 - \frac{2}{\sqrt{n-2}}\sqrt{s_a^2 s_b^2 - cov_{ab}^2}\, t_{n-2,1-\alpha}; +\infty\right),$$

where $cov_{ab} = \frac{\sum_{i=1}^n x_{ia}x_{ib} - \sum_{i=1}^n x_{ia}\sum_{i=1}^n x_{ib}/n}{n-1}$

Parametric CI's for ratios

$(1-2\alpha)$ CI for ratio of means (Ogawa, 1982)

$$CI_{\bar{x}_a/\bar{x}_b} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A},$$

with

$$A = n(n-1)\bar{x}_b^2 - t^2_{n-1,1-2\alpha} \sum_{i=1}^{n} (x_{ib} - \bar{x}_b)^2$$

$$B = -2[n(n-1)\bar{x}_a\bar{x}_b - t^2_{n-1,1-2\alpha} \sum_{i=1}^{n} (x_{ia} - \bar{x}_a)(x_{ib} - \bar{x}_b)]$$

$$C = n(n-1)\bar{x}_a^2 - t^2_{n-1,1-2\alpha} \sum_{i=1}^{n} (x_{ia} - \bar{x}_a)^2$$

and the side condition of significant denominator $n(n-1)\frac{\bar{x}_b^2}{\Sigma(x_{ib} - \bar{x}_b^2)} > t^2_{n-1,1-\alpha}$.

$(1-\alpha)$ lower confidence bound for ratio of variances (Bonett, 2006)

$$CI_{s_a^2/s_b^2} = \left[\frac{s_a^2}{s_b^2}\{g - (g^2 - 1)^{0.5}\}; +\infty\right),$$

where $g = 1 + \{2(1-\hat{\rho}^2)t^2_{n-2,1-\alpha}\}/(n-2)$

This is the inversion of the Pitman-Morgan test.

Nonparametric CI's for ratios (location)

$(1-2\alpha)$ confidence interval for the ratio of means (Bennett, 1965)

Assume: x_{ia}, x_{ib} with $E(x_a) = \mu(>0)$ and $E(x_b) = \kappa\mu$

Wilcoxon signed rank sum statistic $U^+(\kappa) = U^+ = u_1 + u_2 + \cdots + u_n$,
where $u_i = i$ if $z_i > 0$, 0 otherwise and $z_i = z_i(\kappa) = x_{ib} - \kappa x_{ia}$

Estimates $CI_\kappa = [\hat{\kappa}^-; \hat{\kappa}^+]$ are values of κ^-, κ^+ for which $U^+(\kappa)$ is closest to C_{k^-}, C_{k^+}:

$$|U^+(\hat{\kappa}^-) - C_{\kappa^-}| = \inf_\kappa |U^+(\kappa) - C_{\kappa^-}|, \quad |U^+(\hat{\kappa}^+) - C_{\kappa^+}| = \inf_\kappa |U^+(\kappa) - C_{\kappa^+}|,$$

where C_{κ^-} and C_{κ^+} are computed as $C_{\kappa^-}, C_{\kappa^+} \approx \frac{n(n+1)}{4} \pm z_{1-2\alpha} \left\{ \frac{n(n+1)(2n+1)}{24} \right\}^{0.5}$

$\rightarrow$ exact approach exists

Nonparametric CI's for ratios (scale)

$(1-\alpha)$ lower confidence bound for the ratio of the mean absolute deviation of median (Bonett and Seier, 2003)

- Estimate is ratio of mean absolute deviation from the median (MAD) of samples. MAD of sample j is defined as: $\hat{\tau}_j = \sum_{i=1}^{n} |x_{ij} - \tilde{x}_j|/n$
- Lower (1-α) confidence bound for large samples:

$$CI_{\tau_a/\tau_b} = \left[\exp([\log(\frac{\hat{\tau}_a}{\hat{\tau}_b}) - z_\alpha \{var[\log(\hat{\tau}_a/\hat{\tau}_b)]\}^{0.5}]); +\infty\right)$$

- $var[\log(\hat{\tau}_j)] = (s_j^2/\hat{\tau}_j^2 + ((\bar{x}_j - \tilde{x}_j)/\hat{\tau}_j)^2 - 1)/n$
- $var[\log(\hat{\tau}_a/\hat{\tau}_b)] =$ $var[\log(\hat{\tau}_a)] + var[\log(\hat{\tau}_b)] - 2\hat{\rho}_d\{var[\log(\hat{\tau}_a)]var[\log(\hat{\tau}_b)]\}^{0.5}$
- Pearson correlation $\hat{\rho}_d$ based on squared deviation scores $d_{ia} = |x_{ia} - \tilde{x}_a|$ and $d_{ib} = |x_{ib} - \tilde{x}_b|$

Example for location-scale approach

Lipase example with $n = 51$ paired observations
Transfer successful, if trueness $<$ 10% and receiver's precision is non-inferior to a margin of 1/1.1)
The Bland-Altman plot shows no evidence for a systematic proportional difference

Intervals for the difference (sender-receiver)

	location		scale	
type	estimate	95% confidence limits	point estimate	95% lower limit
param.	1,709.3	(673.0; 2,745.7)	104,504,740	(53,403,531; $+\infty$)

Intervals for the ratio (sender/receiver)

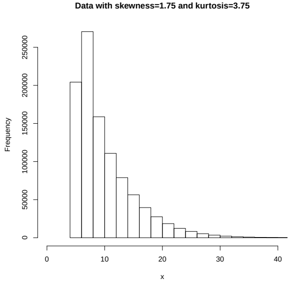
	location		scale	
type	estimate	95% confidence limits	point estimate	95% lower limit
param.	1.027	(1.011; 1.042)	1.075	(1.046; $+\infty$)
nonparam.	1.019	(1.007; 1.033)	1.081	(1.028; $+\infty$)

- a-priori or even a-posteriori definition of absolute scale-variant thresholds of similarity is difficult (especially for precision)
- trueness less than 5% (param ratio: 1.042) and receiver is more precise (nonparam. ratio 1.028)
- transfer is successful

Simulation of the IUT: Set-up

simulated data

- bivariate normal or correlated, nonnormal data with skewness=1.75 and kurtosis=3.75
- $\mu_a = \mu_b = 10$ ($\mu_b = 11.5$ for α)
- $\sigma_a^2 = \sigma_b^2 = 25$ ($\sigma_b^2 = 28.75$ for α)
- $n = 50$ or $n = 500$
- varying correlation from 0 to 0.9



other parameters

- equivalence bounds: $\delta = 0.15, \varepsilon = -0.15, \theta = 1.15, \lambda = 1/1.15$
- $\alpha = 0.05$
- simulations runs: 1,000 for power, 10,000 for α (except nonparametric approach)

Simulation: type I error of IUT

bivariate normal

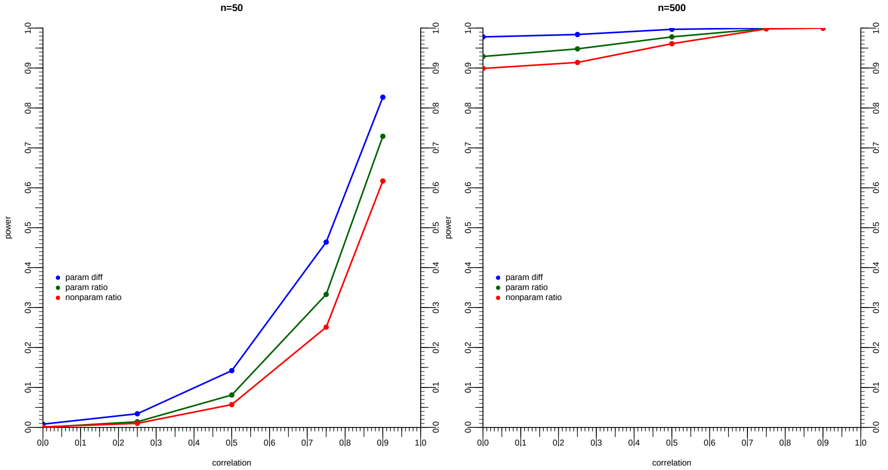
- paired t-test: marginally liberal when $\rho > 0$ in combination with large n

correlated nonnormal

parametric approaches

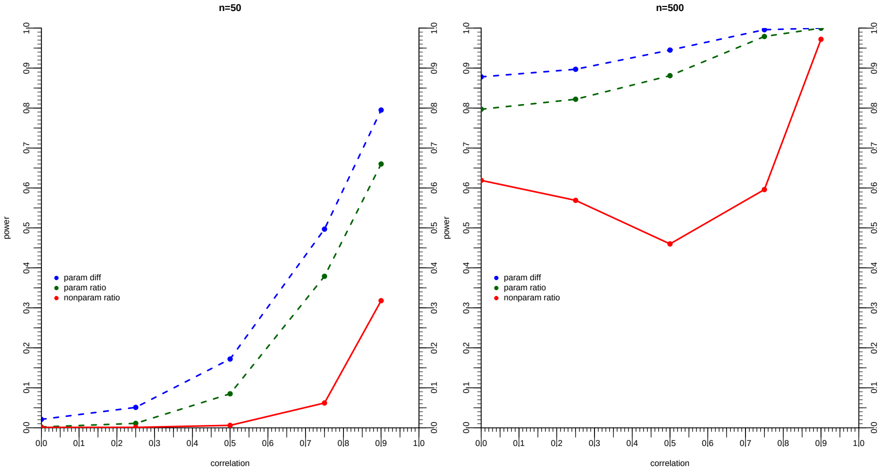
- both location and scale tests:
  increasing α up to 11%-20% with decreasing n and $|\rho|$
- marginal confidence intervals are biased
- IUT still not liberal

Simulation of IUT: normal distributed data



Simulation of IUT: nonnormal distributed data

skewness=1.75 and kurtosis=3.75



Summary and conclusions

- Bland-Altman plot should be routinely used, here for diagnostic of systematic proportional difference
- Bland-Altman plots can be extended with parametric/nonparametric tolerance intervals
- Similarity should be transferred to equivalence for location and non-inferiority for scale
- Margins are assay-specific; if margins cannot be achieved a-priori, post-hoc interpretation is possible with marginal confidence intervals
- Ratios are interpretable and approaches based on ratios are available
- The power of the IUT decreases for smaller variance thresholds more than for smaller location thresholds, i.e. commonly we have to tolerate larger variance confidence intervals for a transfer
- On the other side: if we tolerate a larger variance confidence interval, then the nonparametric test for location can be biased
- R-Code is available on request - a package is planned

References

Altman, D.G. (1990): *Practical Statistics for Medical Research*. Chapman & Hall, 1st ed.

Bennett, B.M. (1965): Confidence Limits for a Ratio Using Wilcoxon's Signed Rank Test. *Biometrics*, 21, 231-234.

Bland, J.M., Altman, D.G. (1986): Statistical method for assessing agreement between two methods of clinical measurement. *The Lancet*, 1 (8476), 307-310.

Bonett, D.G. (2006): Confidence interval for a ratio of variance in bivariate nonnormal distributions. *Journal of Statistical Computation and Simulation* 76 (7), 637-644.

Bonett, D G. and E. Seier (2003): Statistical Inference for a Ratio of Dispersions Using Paired Samples. *Journal of Educational and Behavioral Statistics*, 28 (1), 21-30.

Bradley, E. and L.G. Blackwood (1989): Comparing Paired Data: A Simultaneous Test for Means and Variances. *The American Statistician*, 43 (4), 234-235.

Carstensen, B. (2010): *Comparing Clinical Measurement Methods*, Wiley, 1st ed.

Choudhary, P.K., Nagaraja, H.N. (2005): Assessment of Agreement Using Intersection-Union Principle. *Biometrical Journal*, 47(4), 1-8.

Guidance for Industry: Protocols for the Conduct of Method Transfer Studies for Type C Medicated Feed Assay Methods, FDA, Center for Veterinary Medicine, April 2007.

Guidance for Industry: Bioanalytical Method Validation, US Department of Health and Human Services, US Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Rockville, May 2001.

Guideline on validation of bioanalytical methods, Committee for medicinal products for human use (CHMP), Doc.Ref: EMEA/CHMP/EWP/192217/2009, London, November 2009.

Hahn, G.J. and W.Q. Meeker (1991): *Statistical Intervals: a Guide for Practitioners*, Wiley, 2nd ed.

Hoffman, D. and R. Kringle (2007): A Total Error Approach for the Validation of Quantitative Analytical Methods. *Pharmaceutical Research*, 24 (6), 1157-1164.

Ogawa, J. (1983): Ann. Inst. Statist. Math. 35 (1983) Part A 41.

Rozet, E., Dewé, W., Ziemons, E., Bouklouze, A., Boulanger, B. and Ph. Hubert (2009): Methodologies for the transfer of analytical methods: A review. *Journal of Chromatography B*, 877, 2214-2223.

Zhong J., K. Lee, Y. Tsong, J. Biopharm. Stat. 18 (2008) 1005.

Thank you very much for your attention!