Statistical methods for equivalency test in oxymetry comparison study

YUN BAI[1], Zengri Wang[1], Ted Lystig[1], and Baolin Wu[2]

[1]Medtronic plc; [2]University of Minnesota

Introduction

- Motivation

- An oxymetry comparison study was conducted to demonstrate equivalency in performance between an FDA-cleared oxymetry system and an investigational system in a cohort of healthy, non-smoking adults and adolescent volunteers.
- The study obtained multiple paired measures per individual and the difference of tissue oxygen saturation levels were denoted as Y_{ij} for measure $j = 1, \cdots, m_i$ and subject $i = 1, \cdots, n$. Data were fitted using a random-effects ANOVA model, $Y_{ij} = \mu + u_i + \varepsilon_{ij}$, with $u_i \sim \mathbf{N}(0, \sigma_b^2), \varepsilon_{ij} \sim \mathbf{N}(0, \sigma_w^2)$.
- Per FDA guideline on establishing the equivalency of two oxymetry systems, it is of interest to test the root mean squares, $H_0: \rho \geq \rho_0$ versus $H_a: \rho < \rho_0$, where $\rho = \sqrt{\mu^2 + \sigma_w^2 + \sigma_b^2}$. For example, $\rho_0 = 3\%$ is often used for transmittance, wrap and clip pulse sensors.

- Problems

- ρ is a composite measure involving mean and variance parameters, and poses significant challenges for statistical inference.
- Existing methods are either exploratory or relying on large-sample normal approximation, leading to conservative and unsatisfactory performance.

- New Approach

- Use of novel generalized inference principle to test the hypothesis and set confidence interval for the parameter ρ.
- Improve power with very good finite-sample performance.
- Develop public R package for practical use.

Methods

- Generalized Inference

- Construct a generalized pivotal quantity/test statistic (GT)
  - Being a function of parameters $(\mu, \sigma_w^2, \sigma_b^2)$, *observed data* (y_{ij}), and random variables (Y_{ij}).
    - Traditional pivotal statistics involve only Y_{ij} (or y_{ij}).
  - Reducing to ρ at the observed data (when $Y_{ij} = y_{ij}$).
  - Having a known distribution free of any unknown parameters.
- Key results (ALL properties of normal linear regression model)
  - $\bar{Y}_i = \sum_{j=1}^{m_i} Y_{ij}/m_i \sim N(\mu, \sigma_b^2 + \sigma_w^2/m_i)$; $S_i^2 = \sum_{j=1}^{m_i}(Y_{ij} - \bar{Y}_i)^2/(m_i - 1) \sim \sigma_w^2 \chi_{m_i-1}^2/(m_i - 1)$.
  - $SSE = \sum_i (m_i - 1) S_i^2 \sim \sigma_w^2 \chi_{N-n}^2$, where $N = \sum_i m_i$.
  - $SSR = \sum_i W_i (\bar{Y}_i - \bar{Y})^2 \sim \chi_{n-1}^2$, $W_i = \frac{1}{\sigma_b^2 + \sigma_w^2/m_i}$, $\bar{Y} = \frac{\sum_i W_i \bar{Y}_i}{\sum_i W_i}$.
  - $Z = (\bar{Y} - \mu)\sqrt{\sum_i W_i} \sim \mathbf{N}(0, 1)$.
  - (Z, SSE, SSR) are INDEPENDENT! All distributions are EXACT!

- Proposed GT: intuitive ideas

- 3 steps: (1) Given SSE, estimate σ_w^2; (2) Given σ_w^2 and SSR, estimate σ_b^2; (3) Given σ_w^2, σ_b^2 and Z, estimate μ.
- From $SSE/SSR/Z$, which *exactly* follow $\chi_{N-n}^2/\chi_{n-1}^2/\mathbf{N}(0,1)$, to reversely solve $\sigma_w^2/\sigma_b^2/\mu$, and hence ρ^2.
- NO large-sample normal approximation! EXACT distributions!
- Powerful and efficient test and CI calculation!

Results: Oxymetry comparison study

For illustrative purpose, we analyze data from a subset of 16 individuals. Summary data are provided in the following Table. The observed $SSE = 221.037\%^2$.

Individual i	m_i	$\bar{Y}_i$	i	m_i	$\bar{Y}_i$	i	m_i	$\bar{Y}_i$	i	m_i	$\bar{Y}_i$
1	9	-0.026	5	5	-2.587	9	10	0.963	13	2	-4.333
2	10	0.447	6	10	-0.610	10	10	0.643	14	10	-2.807
3	10	0.083	7	10	0.040	11	10	-0.200	15	10	0.563
4	10	-0.103	8	10	-0.593	12	10	-1.337	16	10	-0.797

- Summary data is adequate to perform the test with no need of individual level data.
- Proposed GT: **Significance p-value: 0.006**; 90% CI for ρ: **[1.665, 2.528]%** .
- Large-sample approximate Z-test: **Significance p-value: 0.010**; 90% CI for ρ^2: [-1.447, 7.221]%^{2}, leading to 90% CI for ρ: **[0, 2.687]%**.
- Both tests support the equivalency of the investigational system and the FDA-cleared system. The proposed GT yields stronger evidence with more significant p-value.

Simulation

- Simulation Study Design

We have conducted comprehensive simulations under various scenarios to investigate the performance of proposed methods. Overall the proposed GT performs significantly better than the large-sample Z-test.

Here we show the results for 10^4 simulations with parameters, $\mu = -0.57, \sigma_w = 1.48, \sigma_b = 1.38$, which are estimated from the pulse oxymetry comparison study. We consider $n = 16$ subjects with m_i ranging from 5 to 20, and $n = 20$ subjects with m_i ranging from 5 to 24. We set $\rho_0 = 2.1$ under the null hypothesis and $\rho_0 = 3$ for estimating power.

- Type I Error

The following Table summarizes the type I errors. Overall the type I errors are well controlled for the proposed GT, while the Z-test is more conservative especially at more stringent significance level.

	$n = 16$		$n = 20$	
α	0.05	0.01	0.05	0.01
GT	0.028	0.005	0.032	0.006
Z-test	0.023	0.0003	0.026	0.0004

- Power

The following Table summarizes the power (%). Overall the proposed GT performs much better than the Z-test.

	$n = 16$		$n = 20$	
α	0.05	0.01	0.05	0.01
GT	82.1	49.3	92.7	68.2
Z-test	72.1	3.2	86.2	16.6

Discussion

- Conclusions

- Proposed a novel approach to testing equivalency in paired medical device comparison studies. And provided strong evidence to support system equivalency for the pulse oxymetry equivalency study.
- Improved power compared to conventionally normal approximation method.
- Justified analytically and developed R package for practical use. The novel approach timely bridged an existing gap in the field.

- References

Tsui K. and Weerahandi S. (1989). Pennello G. (2003). Iyer H., Wang C. and Mathew T. (2004). Ndikintum N. and Rao M. (2016). Bai Y., Wang Z., Lystig T. and Wu B. (2019).

- Acknowledgement

Supported in part by NIH grant GM083345 and NSF grant DMS-1902903.