Generalized Pairwise Comparison as a Highly Versatile Approach to the Design and Analysis of Clinical Trials

Ron Yu
Gilead Sciences

Jeetu Ganju
Ganju Clinical Trials, LLC

BBSW
October 24, 2023

Outline

1. GPC method background
2. Applications beyond efficacy composite endpoint
3. Design considerations
4. Analysis methods
5. Future research

GPC Method – Background

Generalized pairwise comparison (GPC) method is based on the Mann-Whitney U test:

Mann and Whitney (1947)

$$U(X_i, Y_j) = U_{ij} = \begin{cases} +1, & \text{if } X_i > Y_j \\ 0, & \text{if } X_i = Y_j \\ -1, & \text{if } X_i < Y_j \end{cases}$$

and test statistic $U = \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}$.

Main publications on generalized pairwise comparison include:

- **Finkelstein and Schoenfeld (1999)**: *Combining mortality and longitudinal measures in clinical trials*
- **Buyse (2010)**: *Generalized pairwise comparisons of prioritized outcomes in the two-sample problem*
- **Pocock, Ariti, Collier and Wang (2012)**: *The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities*

GPC Method - Example

Heart failure trial

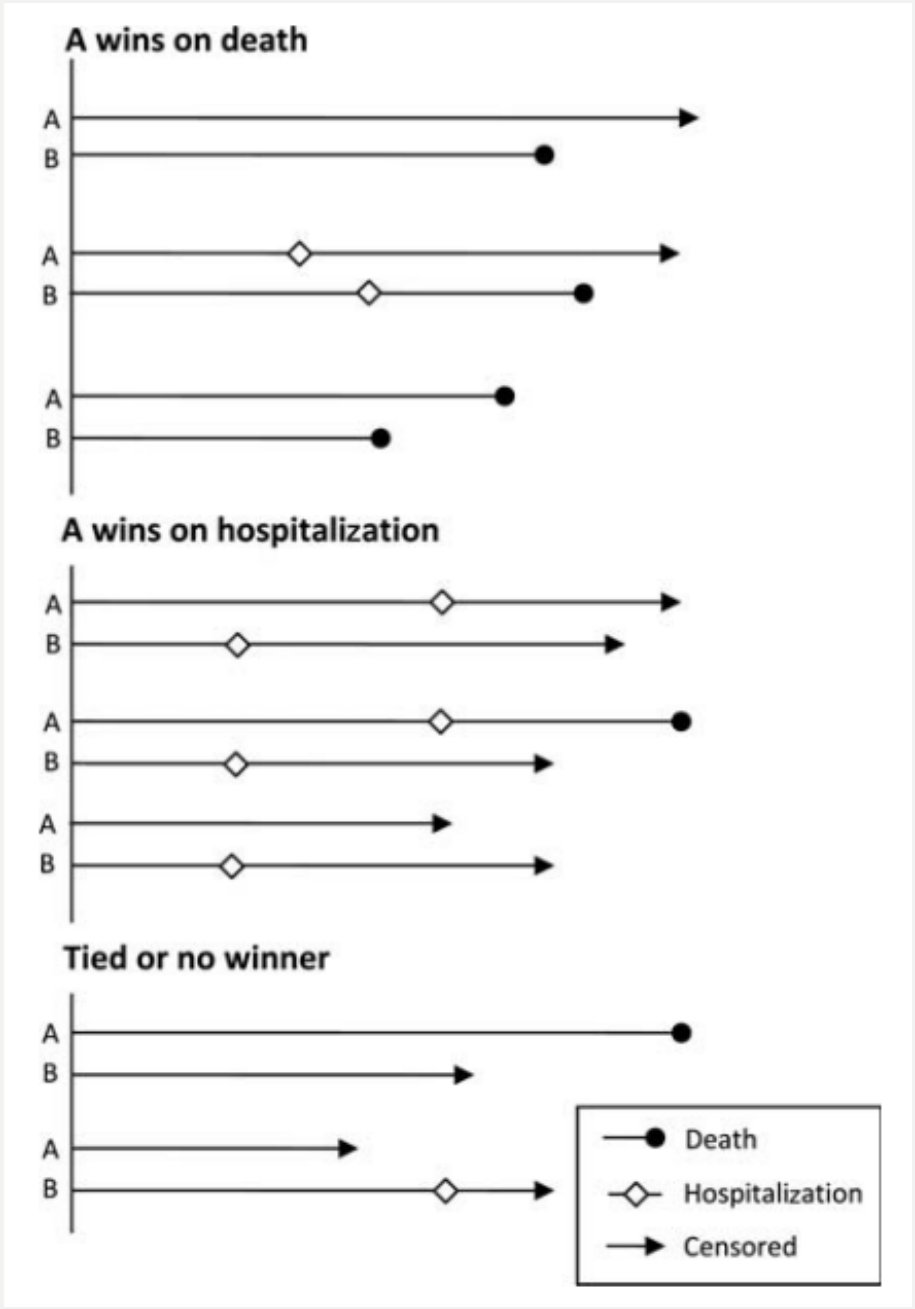
- Time to cardiovascular (CV) death
- Time to heart failure (HF) hospitalization

CV death is more important than HF hospitalization

Two important points

- Each pair is compared at minimum of their follow-up times
- Patients compared on HF hosp. if tied on CV death

Win ratio = $\frac{\#\ \text{wins}}{\#\ \text{losses}}$



(Pocock et al. 2012, Figure 1)

Applications Beyond Efficacy Composite Endpoint

GPC method is simple and flexible, and only relies on a comparison rule

- Between any two selected trial participants, it is a fair question to ask which patient had a more favorable outcome

GPC method can address other important questions

- Benefit-risk quantification
- Missing data handling

(joint work with Arno Fritsch, Katharina Mueller and Patrick Schloemer at Bayer AG)

Benefit-Risk Quantification

Conventional benefit-risk assessment

- Benefits and risks of a medical intervention are evaluated based on aggregated efficacy and safety data separately
- Benefit-risk profile is favorable if benefits outweigh risks – a subjective comparative evaluation

Sequential win ratio analysis

- Pre-specify important **efficacy** and **safety** endpoints and arrange the endpoints by their priorities before study unblinding
- 1st analysis: include the endpoint at the top of the hierarchy
- 2nd analysis: include top two endpoints in the hierarchy
- Etc.

The sequential analyses indicate the extent to which inclusion of endpoints in a stepwise manner alters the assessment of benefit-risk.

Benefit-Risk Quantification: the PATENT-1 Trial

- **Population:** patients with pulmonary arterial hypertension
- **Trial design:** randomized, double-blind, placebo-controlled
- **Primary endpoint:** change from baseline to week 12 in six-minute walk distance
- **Patient follow-up:** 12 weeks
- **# of arms:** 3 (for our analysis we consider one active arm vs placebo)
- Placebo N = 126, Active arm (riociguat 2.5 mg-maximum group) N = 254

For benefit-risk, we propose the following hierarchical **efficacy** and **safety** endpoints

time to death, time to clinical worsening excluding death, time to serious adverse events, and change from baseline to week 12 in 6MWT (minimum distance of 10 meters)

(Ghofrani et al. 2013)

Benefit-Risk Quantification: the PATENT-1 Trial (Result)

2.5 mg-maximum group vs. placebo

Endpoint	Win Ratio	#Wins	#Losses	#Ties (%)	95% CI
Death	**4.29**	489	114	31401 (98.1%)	(0, ∞)[a] (0.34, 53.78)[b]
+ Clinical worsening	**6.32**	1461	231	30312 (94.7%)	(1.38, ∞)[a] (1.42, 28.25)[b]
+ SAE	**1.53**	4517	2948	24539 (76.7%)	(0.81, 2.76)[a] (0.78, 3.02)[b]
+ 6MWT	**1.73**	18702	10797	2505 (7.8%)	(1.34, 2.27)[a] (1.33, 2.26)[b]

[a] based on bootstrap [b] based on variance formula in Yu and Ganju (2022)

Pre-print available on Research Square at https://www.researchsquare.com/article/rs-3221975/v1.pdf

Missing Data Handling

Conventional missing data analysis

- One primary analysis (e.g., mixed model for repeated measures [MMRM])
- One or more sensitivity analyses (e.g., pattern mixture models)

Model assumptions may not hold. Does not distinguish between different reasons for missingness.

Our idea is to incorporate reason for and timing of missingness in pairwise comparisons as a nonparametric approach to handling missing data.

Missing Data Handling: the PATENT-1 Trial

Primary endpoint: change from baseline to week 12 in six-minute walk distance

There are 4 categories of patients:

1. Completers
2. Completed study but missing Week 12 6MWT
3. Missing Week 12 6MWT due to loss to follow-up, non-compliance with study drug, protocol violation or withdrawal by patient
4. Missing Week 12 6MWT due to death or adverse event

Patient A	Patient B	Comparison Rule
Category 1	Category 1	Patients are compared on observed change in 6MWT at Week 12
Category 1	Category 2, 3 or 4	Patient A wins
Category 2	Category 2	Tie
Category 2	Category 3 or 4	Patient A wins
Category 3	Category 3	Tie
Category 3	Category 4	Patient A wins if Patient B discontinues first; otherwise, it is a tie
Category 4	Category 4	Patient whose death or adverse event occurs later wins

Missing Data Handling: the PATENT-1 Trial (Result)

Number and percentage of patients with missing 6MWT at Week 12

Investigator-Reported Reason	Riociguat* (N=254)	Placebo (N=126)
Death	0	2 (1.6%)
AE leading to early withdrawal	8 (3.1%)	6 (4.8%)
Other non-complete	9 (3.5%)	5 (4.0%)
Lost to follow-up	1 (0.4%)	0
Non-compliance with study drug	1 (0.4%)	0
Protocol violation	1 (0.4%)	2 (1.6%)
Withdrawal by subject	6 (2.4%)	3 (2.4%)
Completed study with missing 6MWT at Week 12	4 (1.6%)	1 (0.8%)

* 2.5 mg-maximum group

Result

- #Wins = 19,958 (62%), #Losses = 11,741 (37%), #Ties = 305 (1%)
- Win Ratio = 1.70 and 95% CI = (1.33, 2.22)

Missing Data Handling – Estimand Framework

The method

- Can account for data that are missing not at random (MNAR)
- Is in line with ICH E9(R1) addendum

Composite variable strategies

This relates to the variable of interest (see A.3.3.). An intercurrent event is considered in itself to be informative about the patient's outcome and is therefore incorporated into the definition of the variable. For example, a patient who discontinues treatment because of toxicity may be considered not to have been successfully treated. If the outcome variable was already success or failure, discontinuation of treatment for toxicity would simply be considered another mode of failure. Composite variable strategies do not need to be limited to dichotomous outcomes, however. For example, in a trial measuring physical functioning, a variable might be constructed using outcomes on a continuous scale, with subjects who die being attributed a value reflecting the lack of ability to function. Composite variable strategies can be viewed as

Measure of Treatment Effect

Choice for measure of treatment effect based on GPC

- Net treatment benefit (NTB) $= \frac{\#\text{ Wins} - \#\text{ Losses}}{\#\text{ Pairs}}$
- Win ratio (WR) $= \frac{\#\text{ Wins}}{\#\text{ Losses}}$
- Win odds (WO) $= \frac{\#\text{ Wins} + 0.5 \times \#\text{ Ties}}{\#\text{ Losses} + 0.5 \times \#\text{ Ties}}$

All of the above are referred to as *Win Statistics* (Dong et al. 2021)

Design Consideration – Patient Comparison Rule

- Patient comparison rule can be devised according to
  - Priorities of different outcomes based on clinical judgement, and/or
  - Results of patient preference studies

- Multiple patient comparison rules can be evaluated to test the sensitivity of trial results to different patient preference assumptions

Design Consideration – Sample Size

- In Yu and Ganju (2022),

$$N \approx \frac{\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{\ln^2(WR_{\text{true}})} \times \left\{\frac{4\left(1 + p_{\text{tie}}\right)}{3k(1-k)\left(1 - p_{\text{tie}}\right)}\right\},$$

Assumption:

No intransitivity

where

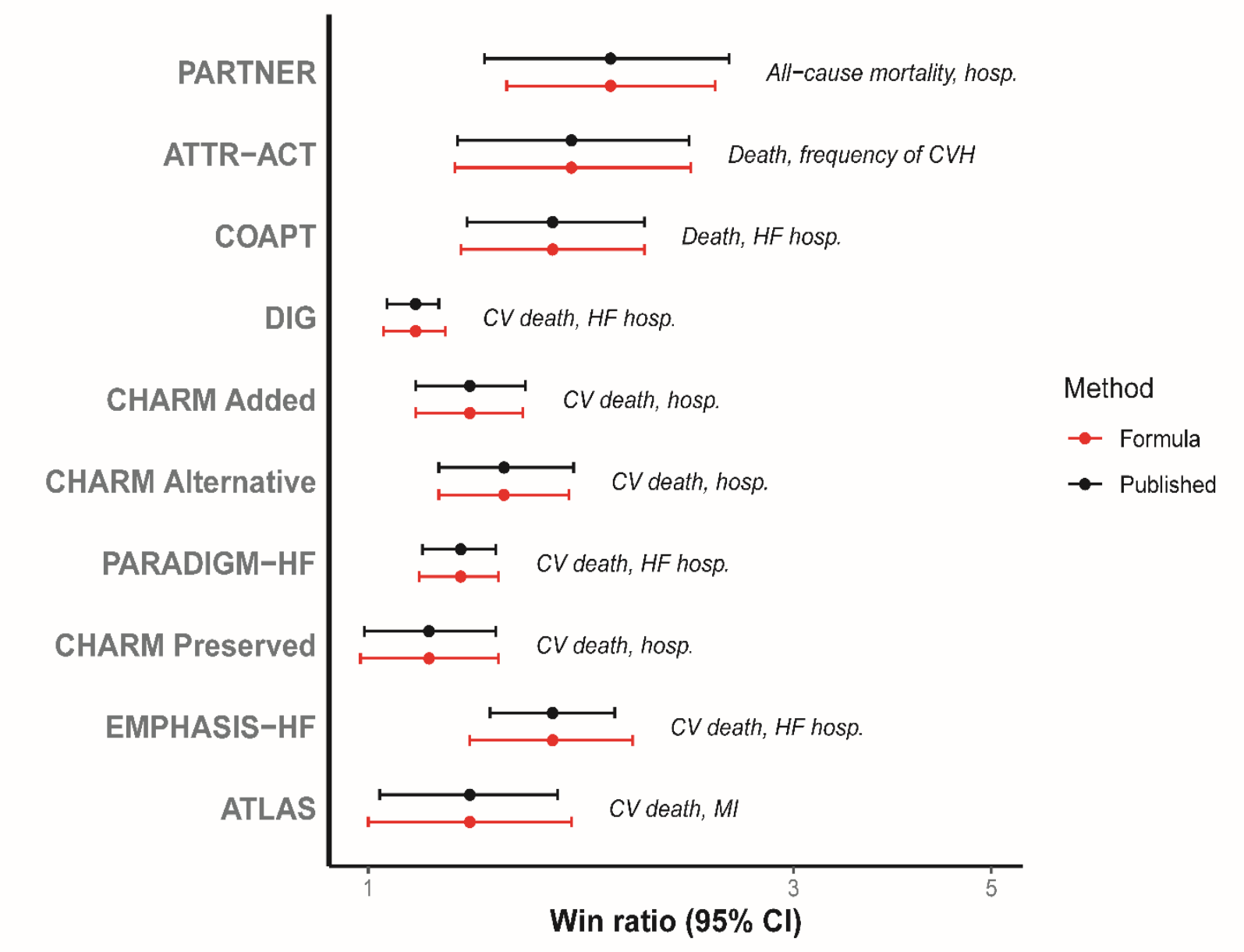
N - total sample size

$Z_{1-\alpha}$ & $Z_{1-\beta}$ - $(1-\alpha)$- and $(1-\beta)$-quantiles of standard normal distribution

WR_{true} - true value of the win ratio

k - proportion of patients allocated to the treatment group

p_{tie} - probability of a tied comparison



(Yu and Ganju 2022, Figure 1)

Published result required individual level data.

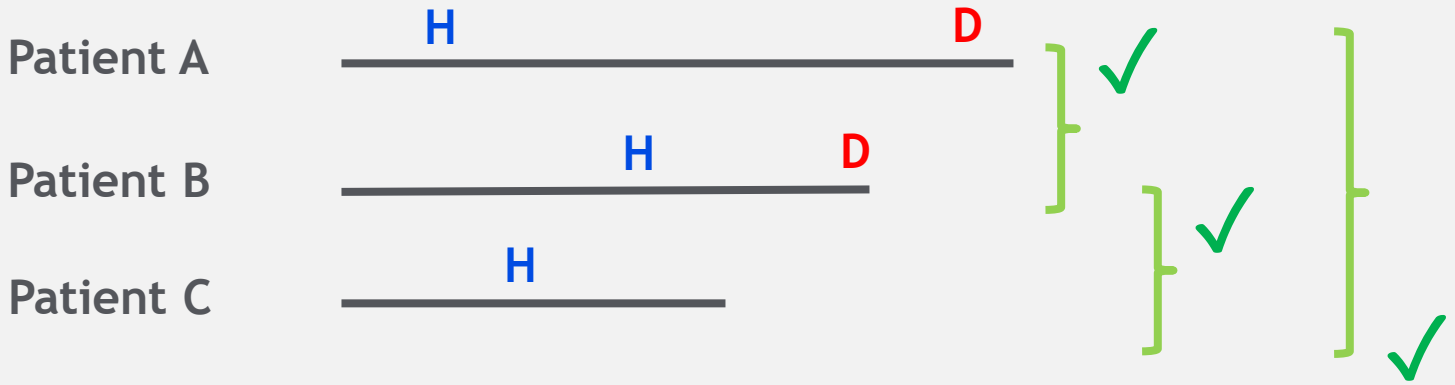
Formula (approximation) requires summary level data

Design Consideration – General Sample Size Formula

- We have developed a general non-parametric sample size formula. This formula holds for any endpoint and allows variable duration of follow-up. It provides a non-parametric answer for endpoints for which sample size formulas are often developed under parametric assumptions – e.g., for a recurrent event endpoint assuming a multiplicative model.

  (joint work with Lu Tian at Stanford University)

This general sample size formula allows intransitivity.



A > B, B > C,

However, A<C!

EP1 - Time to Death EP2 - Time to first Hospitalization

Analysis Methods

Re-sampling methods

- Bootstrap
- Re-randomization test

Normal approximation methods

Choices for variance estimator

- Asymptotic variance for U-statistics
- Exact permutation variance
- Exact bootstrap variance

Future Research – Extension to Recurrent Events

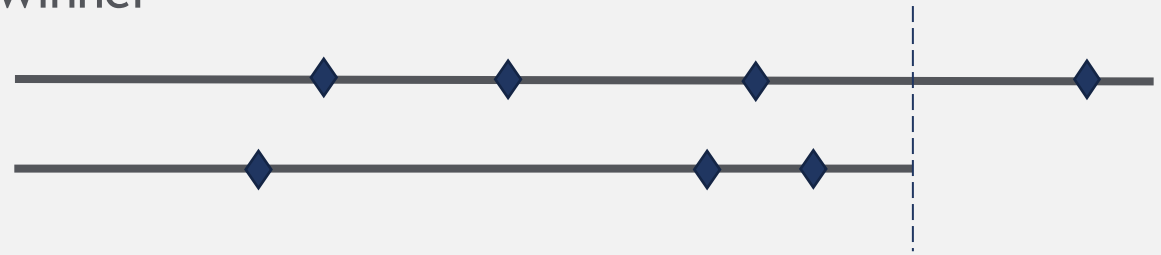
Extension to recurrent events as a nonparametric analysis

- Patients are compared over their shared follow-up time
- The patient with fewer number of events is the winner

Patient A wins

Patient B

- If both patients have the same number of events, time to the 1st event, time to the 2nd event, etc. are compared between the two patients, the patient with a later time to event overall is the winner



Patient A

Patient B wins on time to 2nd and 3rd events, and is the overall winner

Future Research - Extending the Logrank Test

Logrank test statistic for the analysis of time-to-event data:

$$Z = \frac{\sum_{j=1}^{J}(O_j - E_j)}{\sqrt{\sum_{j=1}^{J} V_j}},$$

where O_j and E_j are the observed and expected numbers of events in the treatment group at the j^{th} distinct event time, and $V_j = \mathrm{Var}(O_j - E_j)$

- Limited to the analysis of time-to-event data

Future Research – Logrank-Type Test based on GPC

- Re-expressing the numerator of logrank test as weighted pairwise comparison results:

$$\sum_{j=1}^{J}(O_j - E_j) = \sum_{i=1}^{n}\sum_{k=1}^{m}\frac{I(X_i < Y_k)}{\text{total \# patients at risk at time } X_i} - \sum_{i=1}^{n}\sum_{k=1}^{m}\frac{I(X_i > Y_k)}{\text{total \# patients at risk at time } Y_k},$$

  where $X_1, \ldots, X_n$ are the observations in the treatment group, $Y_1, \ldots, Y_m$ are the observations in the control group, and $I(\cdot)$ is the indicator function

- The above expression can be easily adapted by **GPC method** where the denominators represent the total number of patients in the combined sample who have a follow-up time that is equal to or greater than the follow-up time of the losing patient in a pairwise comparison

Conclusion

- Generalized pairwise comparison (GPC) is a highly versatile approach to the design and analysis of clinical trials
- Although GPC was originally proposed in the context of composite endpoint for efficacy analysis, the method can be used to address other important questions, such as quantifying benefit-risk and handling missing data
- There are many possible ways to extend the methodology which are being examined

References

- Buyse, M. (2010). Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Statistics in Medicine, 29*(30), 3245-3257.
- Dong, G., Huang, B., Wang, D., Verbeeck, J., Wang, J., & Hoaglin, D. C. (2021). Adjusting win statistics for dependent censoring. *Pharmaceutical Statistics, 20*(3), 440-450.
- Finkelstein, D. M., & Schoenfeld, D. A. (1999). Combining mortality and longitudinal measures in clinical trials. *Statistics in Medicine, 18*(11), 1341-1354.
- Ganju, J., Fritsch, A., Mueller, K., Schloemer, P., & Yu, R. (2023). Quantifying Benefit-Risk and Addressing Missing Data in Late-Stage Clinical Trials. *https://www.researchsquare.com/article/rs-3221975/v1.pdf*
- Ghofrani, H. A., Galiè, N., Grimminger, F., Grünig, E., Humbert, M., Jing, Z. C., Keogh, A. M., Langleben, D., Kilama, M. O., Fritsch, A., Neuser, D. & Rubin, L. J. (2013). Riociguat for the treatment of pulmonary arterial hypertension. *New England Journal of Medicine, 369*(4), 330-340.
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 50-60.
- Pocock, S. J., Ariti, C. A., Collier, T. J., & Wang, D. (2012). The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European Heart Journal, 33*(2), 176-182.
- Yu, R. X., & Ganju, J. (2022). Sample size formula for a win ratio endpoint. *Statistics in Medicine, 41*(6), 950-963.
- Yu, R. X., Ganju, J. & Tian L. A Universal Sample Size Formula for Comparison of Two Groups. *To be submitted*.


THANK YOU