Comparison of Statistical Analysis Methods Using Modeling and Simulation for Modern Protocol Design

BASS XIII Meeting

Dejun Tang, Jose Pinheiro,
Chunlin Qian, and Heinz Schmidli
Novartis Pharmaceuticals

Outline

- Introduction
- Background
- Statistical methods and assumptions
- Issues and concerns
- Modeling and simulation
- Simulation results and conclusions
- Summary and discussion

Introduction

- Statistical analysis methods/models usually require certain distributional assumptions.
- In modern protocol design of clinical studies, complex models are widely used. Often times it is difficult to examine the validity of the underlying model assumptions.
- When more than one possible method is available, determining the “optimal” method is also needed.
- This talk discusses and illustrates the use of modeling and simulation to compare alternative statistical methods to be used in a Phase III study.

Background

- FTY720 is an oral investigational drug that is being developed in Novartis as treatment for Multiple Sclerosis (MS) patients.
- Promising Phase II data led to the development and initiation of a Phase III clinical trial program.

Efficacy endpoint

- In Phase III MS studies, an important efficacy endpoint is the annualized relapse rate (ARR), which is defined as the average number of relapses per year.
- Because the number of relapses is a count variable, Poisson regression modeling was initially considered to analyze the relapse data.

Assumptions of Poisson (1)

- Mean parameter λ is a constant over time.
  - This can be considered true for the placebo group.
  - For the FTY720 group, however, the purpose of treatment is to reduce the relapse rate over time. It is expected to observe more relapses in the first year than in the second year.
  - Nevertheless, we can assume we compare the average relapse rates between two groups. The violation to this assumption may not cause severe problems.

Assumptions of Poisson (2)

- Occurrence of events is independent.
  - In a Poisson process, if an event happens at time T_1, then the probability the next event occurs in $(T_1, T_1+ \delta]$ is independent of T_1.
  - However this does not hold for MS relapses, because of the following:
    - A relapse has a duration, which may last up to 90 days.
    - A relapse can not be a confirmed relapse if it is within 30 days of previous confirmed relapse. It means that relapses have a 30-day waiting period.

Assumptions of Poisson (3)

- The variance is equal to the mean λ.
  - This is a key assumption for the Poisson distribution which sometimes is not satisfied in practice.
  - Often, the observed sample variance is larger than the sample mean, consistent with the so-called "overdispersion."
  - The overdispersion factor is defined as δ*100% when variance = (1+δ) mean.

Overdispersion in Phase II study

	Treatment	Overdispersion
6-month data	FTY720-1	34%
	FTY720-2	21%
	Placebo	5%
12-month data	FTY720-1	223%
	FTY720-2	64%

Do we think too much?

One may ask why you are concerned about these things so much:

- Any model is an approximation to the real world and nothing will fit it exactly.
- It is a common practice to use Poisson regression to handle count data.
- If overdispersion is observed, Generalized Estimating Equations (GEE) can be used to adjust the estimates.

Need more explorations

- It is known that the Poisson assumptions are violated. Is Poisson regression (with or without GEE) still a valid statistical method to analyze the data?
- Due to the mix of violations for several assumptions, What is the impact on the power? Should we adjust the sample size?
- To quantify the impact and find the best choice, a simulation study was proposed.

What is the best?

- Valid and most powerful statistical method and efficacy endpoint combination.
- Two criteria must be met:
  - **Valid**: preserves the α-level under the null hypothesis of no treatment difference.
  - **Most powerful**: provides the most power under alternative hypothesis.

Objectives of simulation study

- To evaluate if selected models/methods can preserve the α-level under the null hypothesis
- To estimate the power associated with the calculated sample size
- To examine the impact of dropouts
- To assess problems overdispersion may cause in the methods performance

Statistical methods to be compared

- Poisson regression
- Poisson regression with GEE (overdispersion)
- Andersen-Gill marginal model (original)
- Rank Analysis of Covariance*
- Rank Analysis of Covariance with imputation
- Logistic regression (proportion of relapse-free)
- Cox proportional hazard regression (time to first relapse)

---

* Stokes, Davis, and Koch (2000)

Imputation for Rank ANCOVA

- The following imputation applies to dropout patients:
  - Calculate the monthly mean number of relapses for all patients contributing data to that month, regardless of the treatment.
  - For patients who drop out early, impute the missing data month by month using the monthly mean number of relapses.
- This imputation adds a constant number of relapses to dropout patients at the same month. It is acceptable for rank based analysis methods.

Simulation plan

- Two types of simulation were implemented:
  - Sub-sampling from Phase II data

    Empirically choose a smaller sample size to sub-sample the data and make the p-value for Poisson regression to be 0.05. Then repeat it N times to compare the p-values for all seven methods.
  - Pseudo-random number simulation

    Generate data according to a parametric model closely describing the real relapse process. Then evaluate and compare the analysis methods.

Sub-sampling setting

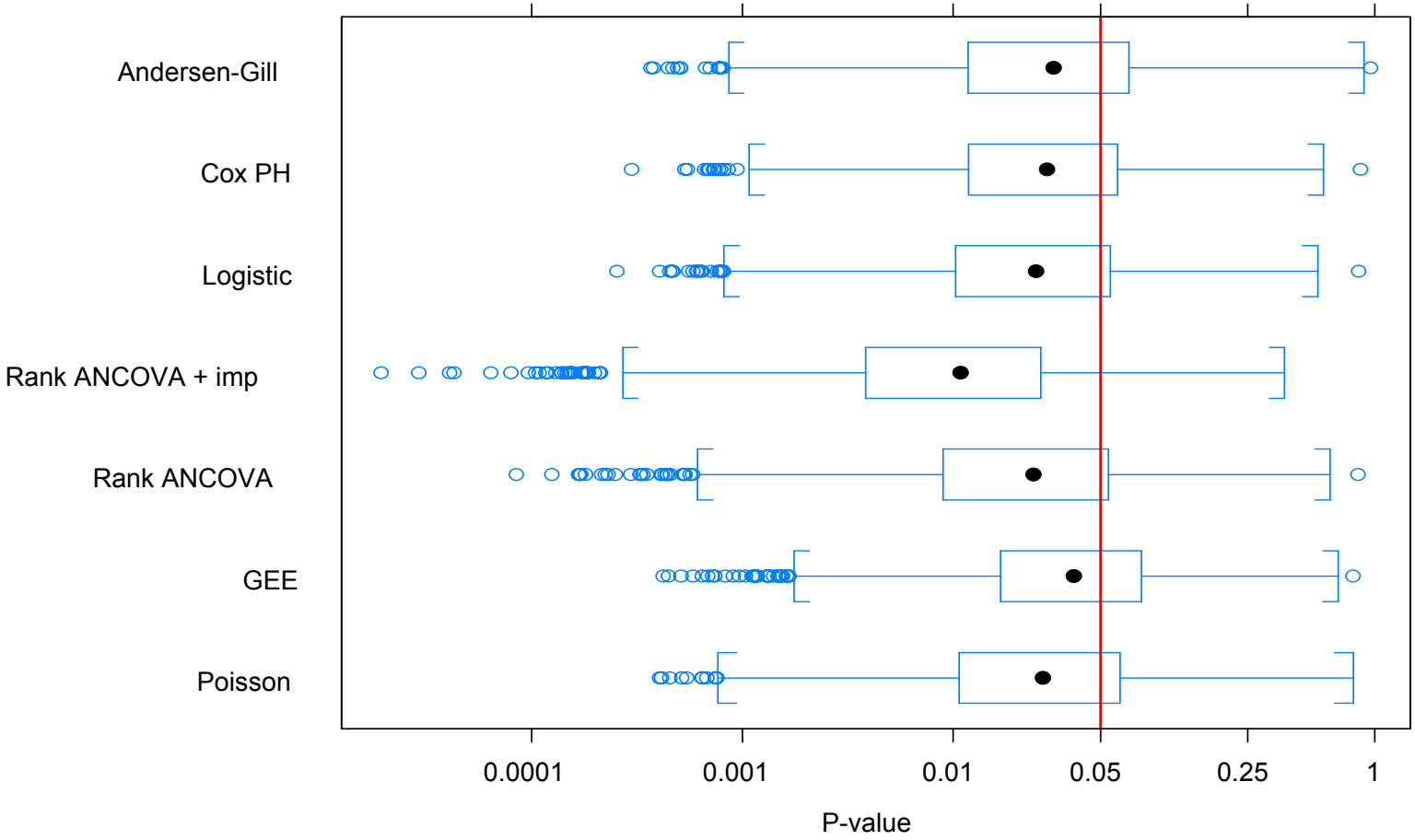
- The original sample size for the Phase II study is 93 patients per arm.
- To make the p-value for Poisson regression be around 0.05, empirically chosen sample size for the sub-sampling is 65 patients per arm.
- Re-sampling size is N=2000.

Sub-sampling results

Statistical Method	p-value (mean)
Poisson	0.0503
Poisson with GEE	0.0618
A-G	0.0561
Rank ANCOVA	0.0446
Rank ANCOVA w imputation	0.0232
Logistic	0.0453
Cox PH	0.0493

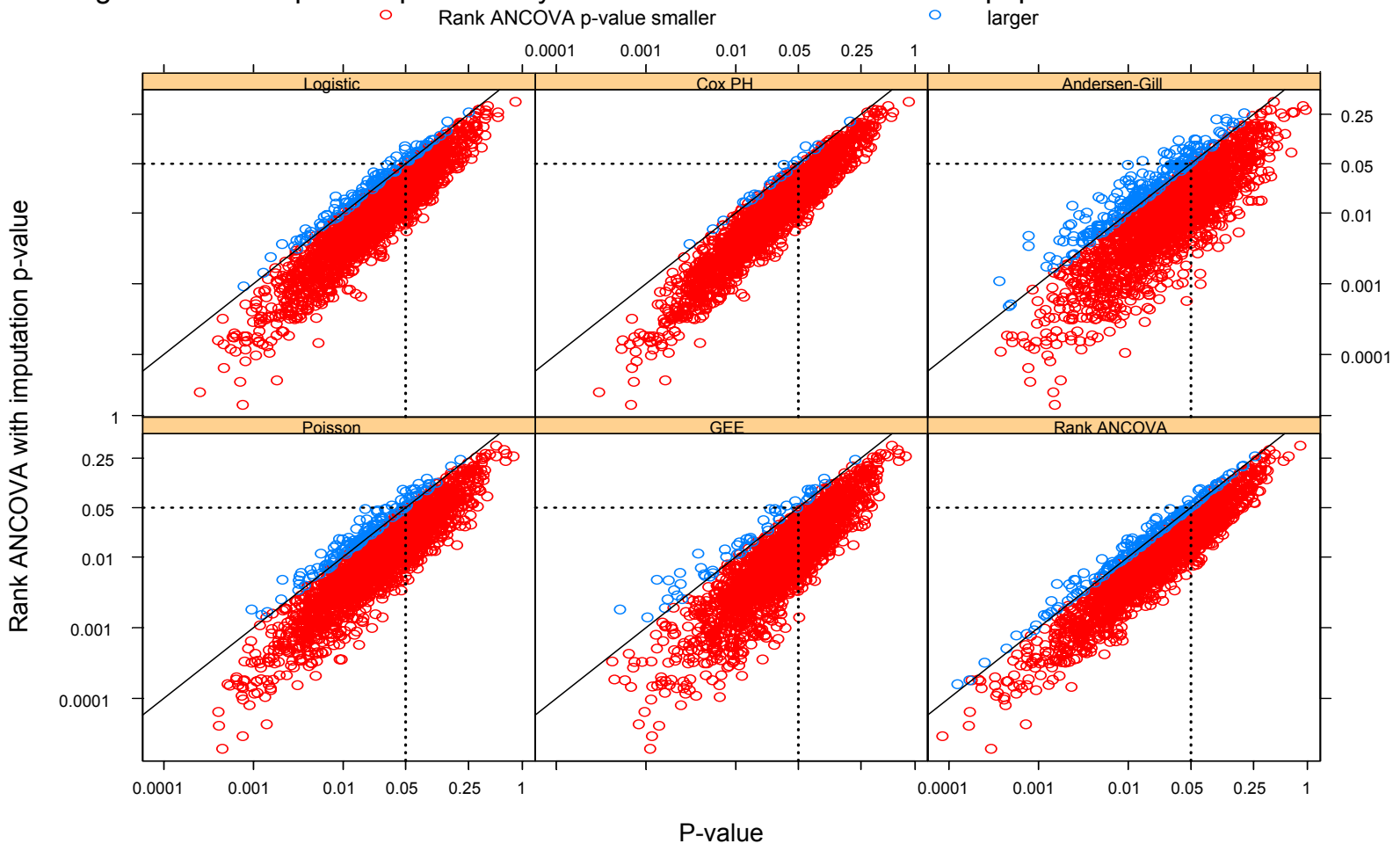
Sub-sampling results (2)

Figure 1: boxplots of subsampling p-values by method



Sub-sampling results (3)

Figure 2: scatter plots of p-values by method vs. Rank ANCOVA + imp. p-values



Pseudo-random simulation setting

- A 24-month study with FTY720 treatment and placebo arms
- Sample size: 367 per arm (calculated based on disability progression)
- Overdispersion: 0% to 125% (by 25%)
- Effect size: 0% to 80% reduction (by 10%)
- Dropout rate: 20%, 25%, and 30% over 24 months
- Simulation size: N=3000+

Key step of Pseudo-random simulation

- To generate the pseudo-random responses reflecting the following:
  - 30-day waiting after onset of relapse
  - Event rate decreasing over time
  - Overdispersion factor
  - Dropouts
- Once we have the data above, the rest is a routine statistical analysis procedure.

Pseudo-random number generation (1)

- 30-day waiting after onset
  - Divide 24 months into disjoint one-month intervals.
  - One month can have at most 1 relapse.
  - If there is no relapse in month i, then one relapse may occur in month i+1.
  - If a relapse occurred in month i at time (i-1)+x, where x is the proportion of time during month i, then a relapse in month i+1 can only occur in (i+x, i+1].

Pseudo-random number generation (2)

- Relapse rate decreases over time
  - Use the following E_{max} model to represent the decreasing relapse rates:

    $\lambda(t) = \lambda^0 - \delta_{max} * t / (ET_{50} + t)$

    where λ^0 is the placebo constant rate, δ_{max} is the maximum asymptotic rate decrease, and ET_{50} is the time till half maximum rate decrease.
  - Model specifies the instantaneous relapse rate at time t. The midpoint of a month interval was used as the relapse rate for that month.

Pseudo-random number generation (3)

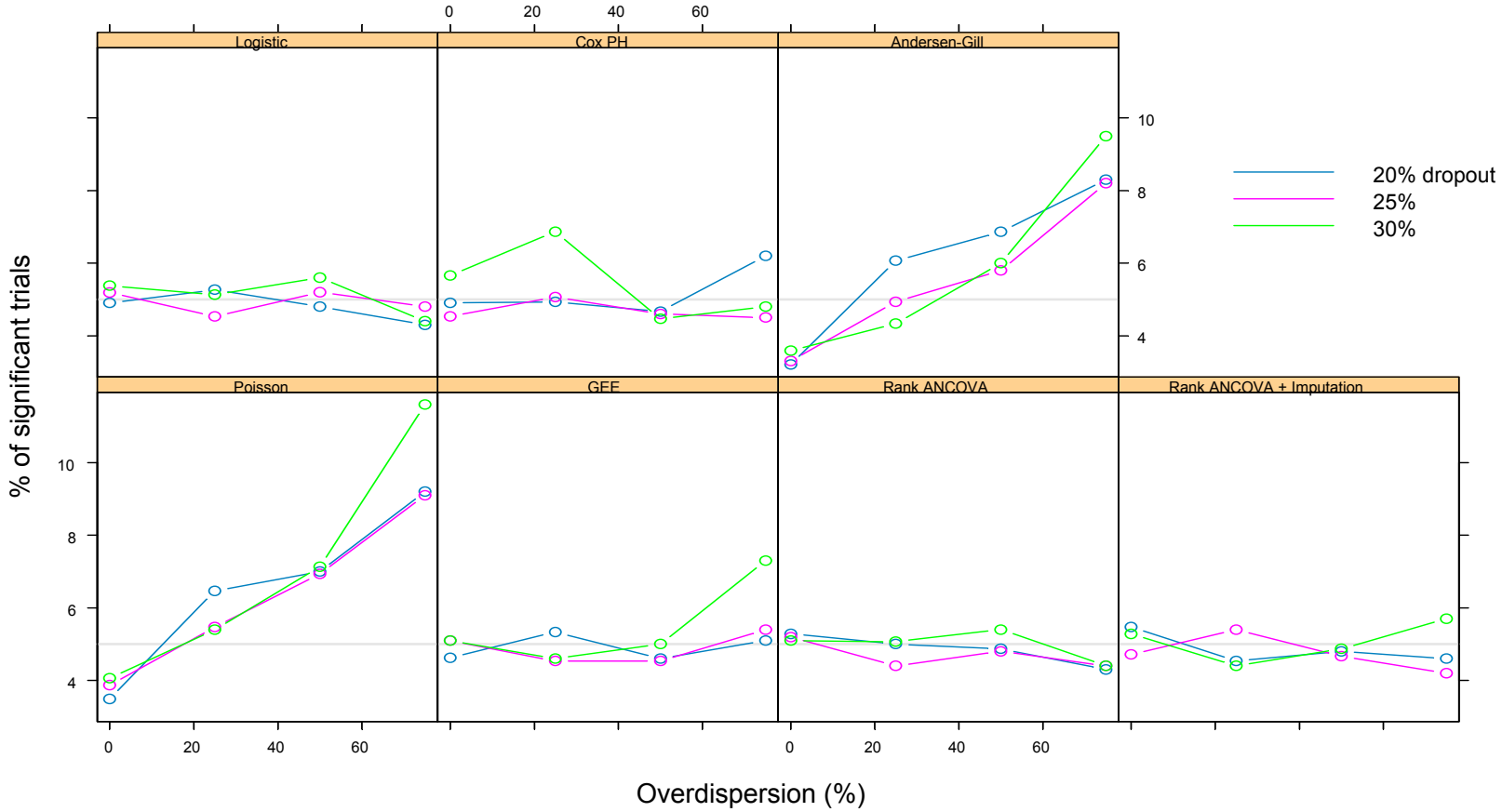
- Overdispersion
  - Conjecture: reason for overdispersion is within-patient correlation.
  - Assume that within-patient relapses follow a homogeneous Poisson process with patient-specific rates $\lambda_j \sim$ log-normally distributed.
  - By adjusting the coefficients of variation (CV) of the log-normal distribution under this hierarchical model, we can determine the overdispersion factor for the relapse rate.

Pseudo-random number generation (4)

- Dropouts
  - Choose a dropout rate (say 20%), simulate the dropout time uniformly over the 24-month time interval for that percentage of patients.
  - The dropout time points for 20%, 25%, and 30% are generated separately.
  - Use the dropout time as the study completion time for these patients.

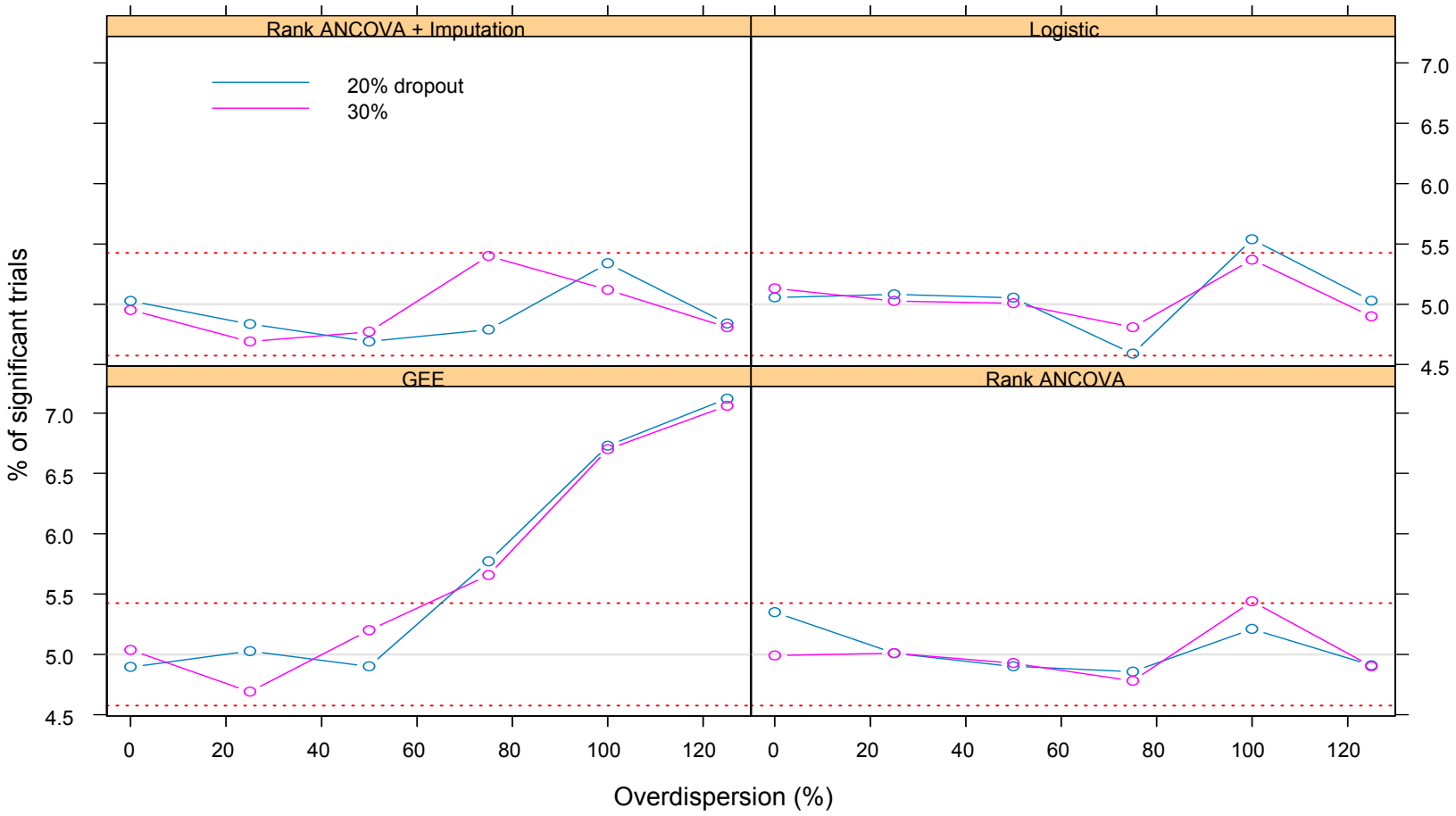
Simulation results - validity (1)

Figure 1: FTY vs. Placebo comparison, 24 month data, alpha = 5%



Simulation results - validity (2)

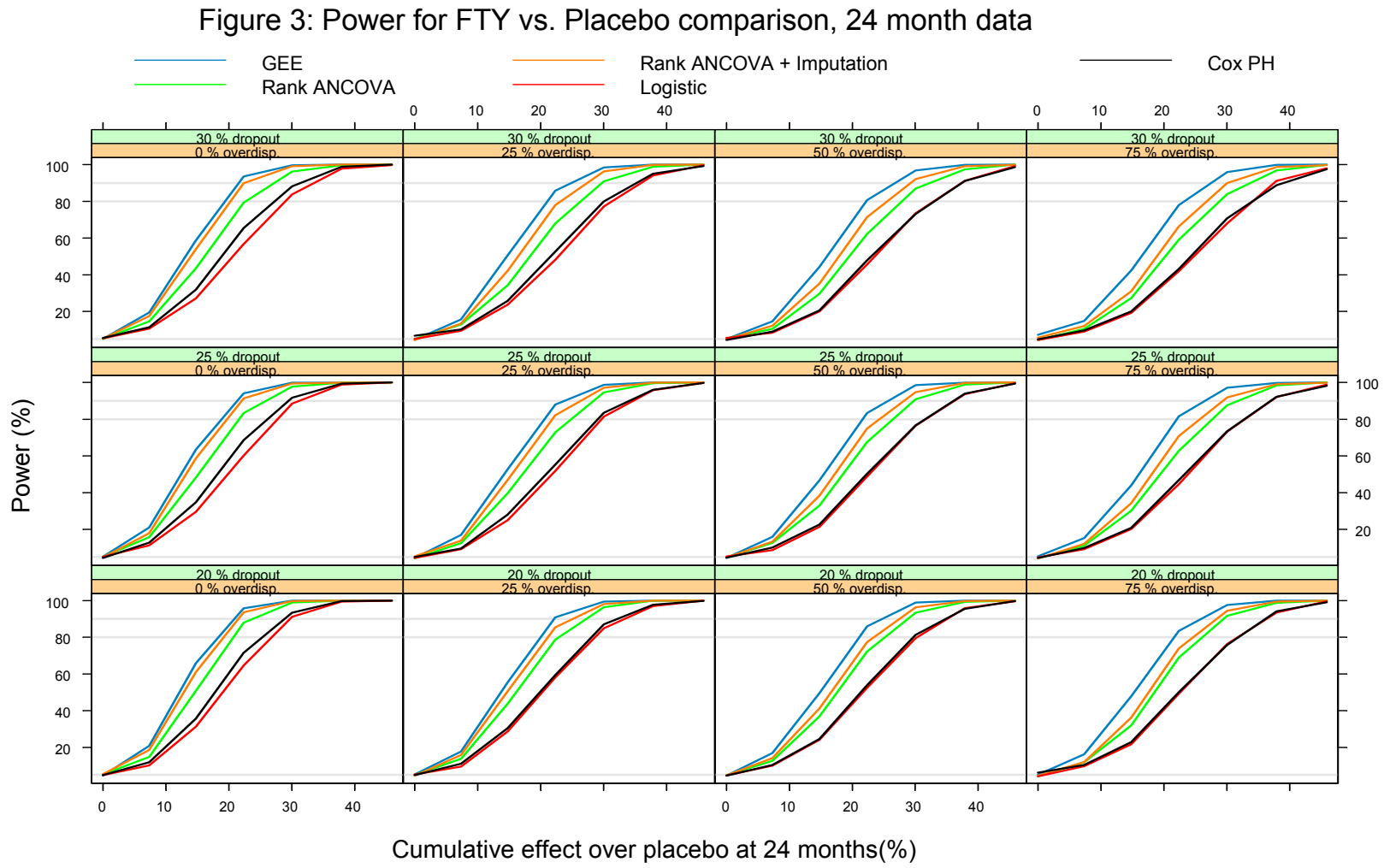
Figure 2: FTY vs. Placebo comparison, 24 month data, alpha = 5%
wider overdispersion range



Simulation conclusions - validity

- Poisson regression and Anderson-Gill method could NOT preserve the α-level under the null hypothesis when overdispersion is >25%.
- Poisson regression with GEE could NOT preserve the α-level either under the null hypothesis when overdispersion is >75%.
- All other methods can retain the α-level in our overdispersion range (<=125%).

Simulation results - power

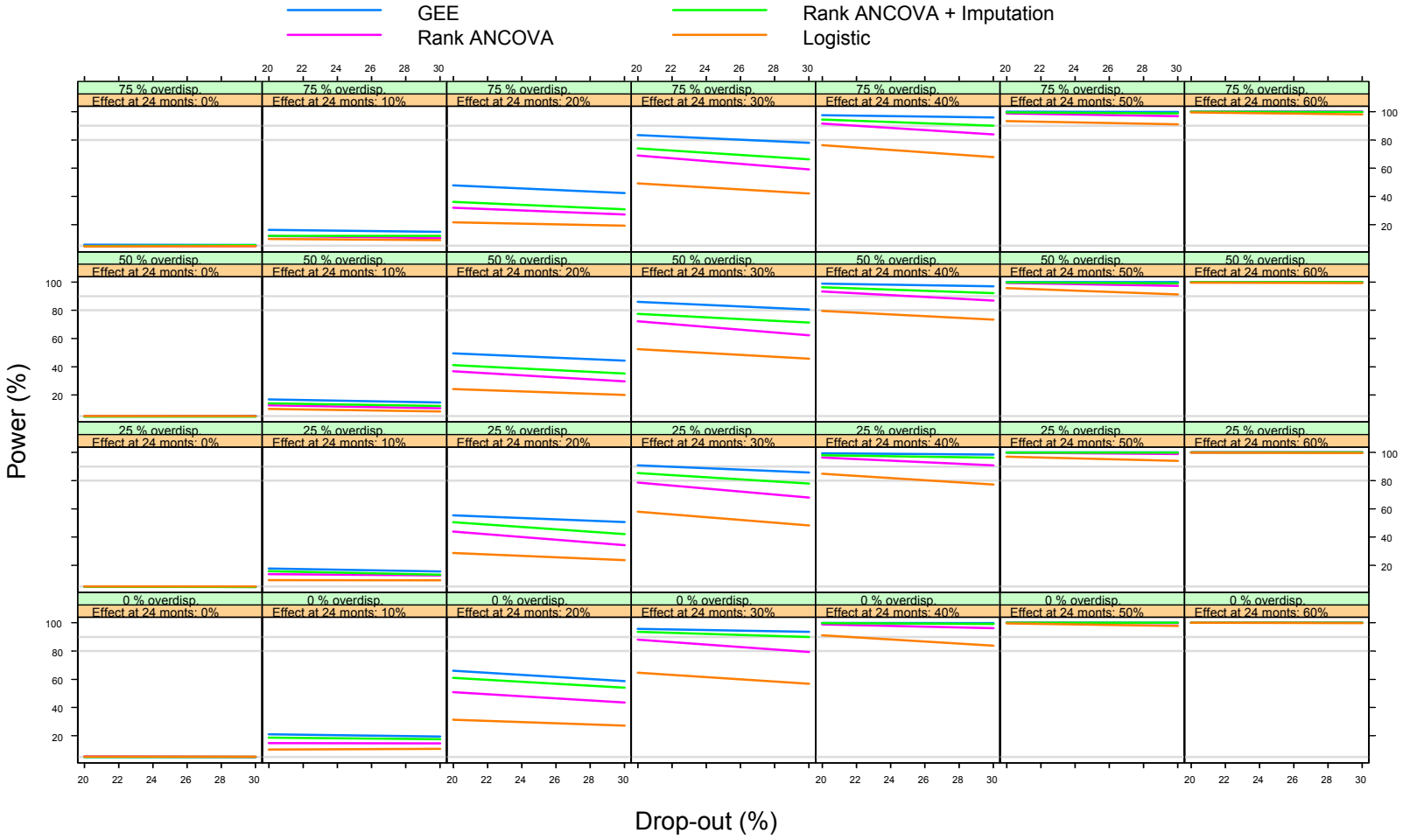


Simulation conclusions - power

- The expected effect size is a relapse rate reduction of 40%.
- From the plots, when reduction of relapse rate is 40% or more, all methods reach the power of 90% or higher for the desired sample size of 367 patients per arm.
- Rank ANCOVA with imputation is the most powerful among valid methods.

Simulation results - dropouts

Figure 7: Power of FTY vs. Placebo comparison as function of drop-out, 24 month data



Simulation conclusion - dropouts

- Higher dropout rate will reduce the power. However, the reduction is very small and can be ignored for practical purposes.
- This is true for all valid statistical models and the overdispersion range we considered.

Summary

- Different endpoints and statistical methods are considered in a protocol design.
- To explore the best choice, modeling and simulation were used.
- Poisson, A-G, and Poisson+GEE are not valid when overdispersion is large (>75%).
- Rank ANCOVA+imputation is the valid and most powerful method for the study.
- Dropout rate has little impact.

Something to explore

- Anderson-Gill method with the empirical variance estimate and removal of 30-day waiting period to assess the performance with larger overdispersion factors.
- Set up a threshold T for the overdispersion factor. If the overdispersion factor > T, then the Rank ANCOVA will be applied; otherwise, GEE will be used, to check if it will improve the power of the test.

Acknowledgment

- We would like to thank the following people for their valuable advice, discussion, and support to this project:
  - Prof. Gary Koch
  - Willi Maurer, Audrey Wong, David Smith, and Tomas Haas at Novartis Pharma
  - Novartis FTY720 MS project team

Questions or comments?

References

[1] Stokes ME, Davis CS, Koch GG (2000), Categorical data analysis using the SAS system. 2nd edition, SAS Institute, Inc. Cary, NC.

[2] Therneau TM, Grambsch PM (2000), Modeling survival data: extending the Cox model, Springer, New York.

[3] Hardin WH, Hilbe JM (2003), Generalized estimating equations, Chapman & Hill, CRC.