Adaptive Operational Testing

Victoria "dB" Sieck

Dr. Fletcher G W Christensen

Outline

- Background and Artificial OT Example
- Moving OT into a Bayesian Framework
- Adaptive OT using Predictive Probability
- Example Results
- Extensions using Informative Priors
- Conclusion

Goal of Proposed Method

- Two overarching acquisition phases: operational testing (OT) and developmental testing (DT)
- Goal: create an efficient and effective OT using Bayesian Methods
  - Analyzing data during testing to support earlier system evaluations
  - Create informative priors for OT using previous testing data (e.g. DT data)
- Research considers the most granular part of system evaluation: evaluating a single question (measure)

Artificial Operational Testing Example

Simulated Example: Electric Semi-Truck

- Transport company procuring an electric semi-truck
- One of the questions (measure) to answer: is the mean number of miles traveled on one charge ≥ 400 miles?
  - Parameter of interest, ϕ, is the mean number of miles
  - Metric threshold that must be obtained, ϕ_0, is 400
- Response variable for design of experiments process: number of miles traveled on one charge

Artificial Example – Factor Management

Factor	Levels	Magnitude of Effect	Likelihood of Encountering
Terrain	Hilly Flat	High	50% 50%
Temperature	Hot (> 70º F) Moderate (70º - 50º F)	Medium	4/9 5/9
Wind	Good Moderate Poor	Medium	1/3 1/3 1/3
Payload Type	Refrigerated Non-Refrigerated	High	50% 50%
Weight	Heavy (≥ 40k lbs) Light(< 40k lbs)	High	50% 50%

Current Operational Testing Process

- Selected an experimental design
  - Main effects and two-way interactions (excluding wind)
  - 2^4 Full Factorial, five replicates
  - Design has 80% power, with 80% confidence
- Test is executed
- Measures are evaluated
  - If $\phi \geq \phi_0$, the measure is evaluated as met
  - If $\phi < \phi_0$, the measure is evaluated as not met

Moving Operational Testing into a Bayesian Framework

THE UNIVERSITY OF NEW MEXICO

A Bayesian Framework for Operational Testing

- Augmenting the current test design process
  - Formalizing the current underlying ANOVA structure (when not explicitly used)
  - Including prior selection to the test design process
- Evaluating a measure: calculate the posterior probability of a system obtaining the metric threshold and compare that to a certainty threshold, $\theta_T = 0.8$:
  - If $\Pr(\phi \geq 400) \geq 0.8$, evaluate the measure as met
  - If $\Pr(\phi \geq 400) < 0.8$, evaluate the measure as not met

Electric Semi-Truck Example: Bayesian Set-Up

THE UNIVERSITY OF NEW MEXICO

- Distribution of miles traveled within groups: $y_{ijklmp}|\mu_{ijklm} \sim N\left(\mu_{ijklm}, \frac{1}{\tau}\right)$, where

$$\mu_{ijklm} = \eta + \alpha_i + \beta_j + \omega_k + \gamma_l + \delta_m + (\alpha\beta)_{ij} + (\alpha\gamma)_{il} + (\alpha\delta)_{im} + (\beta\gamma)_{jl} + (\beta\delta)_{jm} + (\gamma\delta)_{lm}$$

- Reference cell ANOVA model (baseline parameter: η, the first level of each factor)

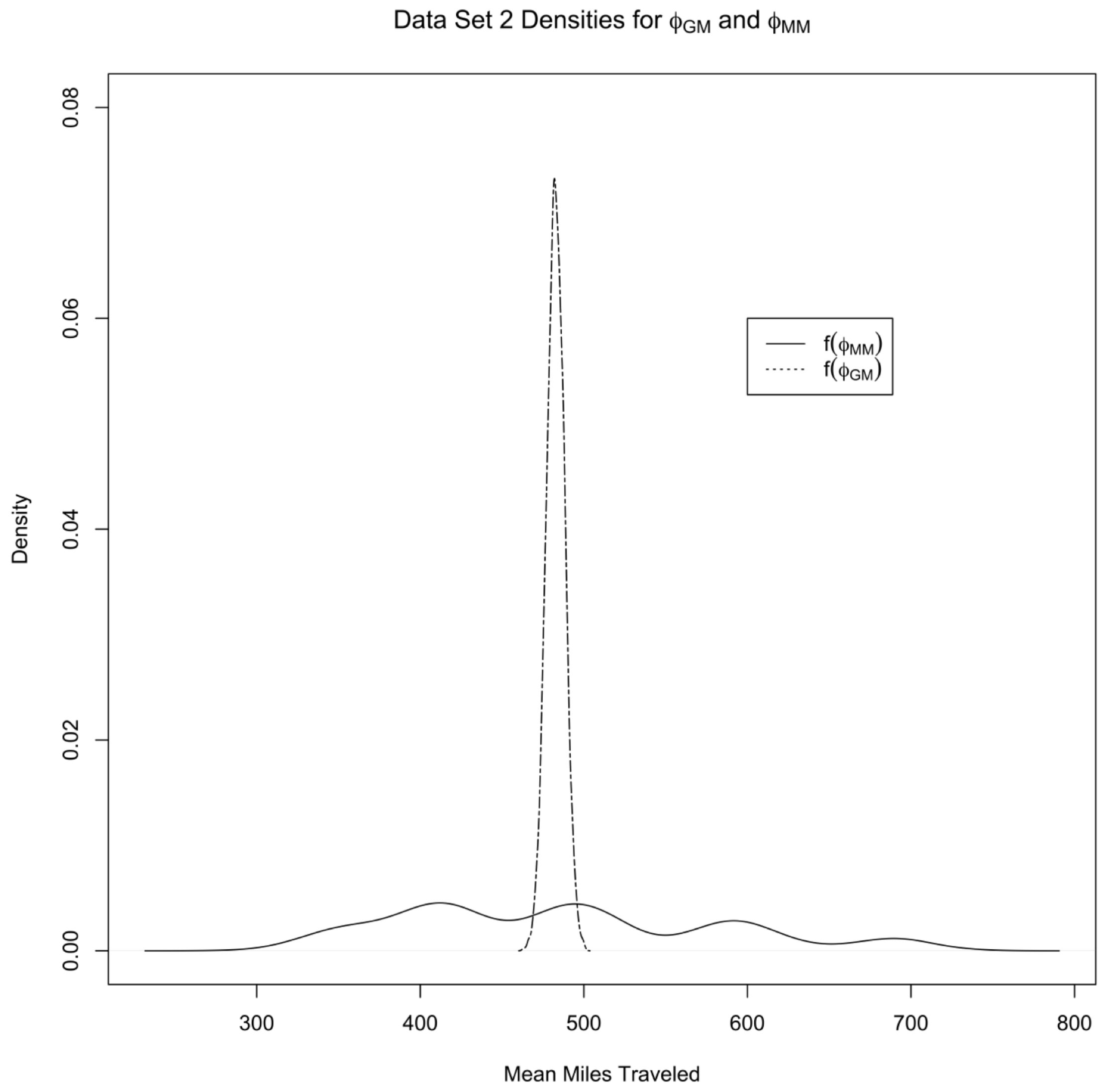
Factor	Levels	Model Parameter	Weakly Informative Priors
Terrain	Hilly Flat	α_i	$\Pr(\alpha_1 = 0)$=1 $p(\alpha_2) \sim N(50, 10000)$
Temperature	Hot (> 70º F) Moderate (70º - 50º F)	β_j	$\Pr(\beta_1 = 0)$=1 $p(\beta_2) \sim N(50, 2500)$
Wind	Good Moderate Poor	ω_k	$\Pr(\omega_1 = 0) = 1$ $p(\omega_2) \sim N(-25, 2500)$ $p(\omega_3) \sim N(-50, 2500)$
Payload Type	Refrigerated Non-Refrigerated	γ_l	$\Pr(\gamma_1 = 0)$=1 $p(\gamma_2) \sim N(100, 10000)$
Weight	Heavy (≥ 40k lbs) Light(< 40k lbs)	δ_m	$\Pr(\delta_1 = 0)$=1 $p(\delta_2) \sim N(100, 10000)$

- 2^4 full factorial, five replicates = 80 test events

Bayesian Mission Mean

- Analysis using an operational focus – “mission sets”
- Mission Sets: Combination of factor levels that represent an operational environment
- Obtaining ϕ, mission mean approach:
  - Sample mission sets
  - Use mission sets and posterior draws induce a distribution on ϕ
  - ϕ is then a mixture distribution of random mission means, μ_{ijklm}
- Marginalize out the mission sets to evaluate the measure
- Bayesian Perspective: grand mean is a weighted average of random variables and mission mean is a random selection of random variables

Bayesian Grand Mean Approach vs Bayesian Mission Mean Approach



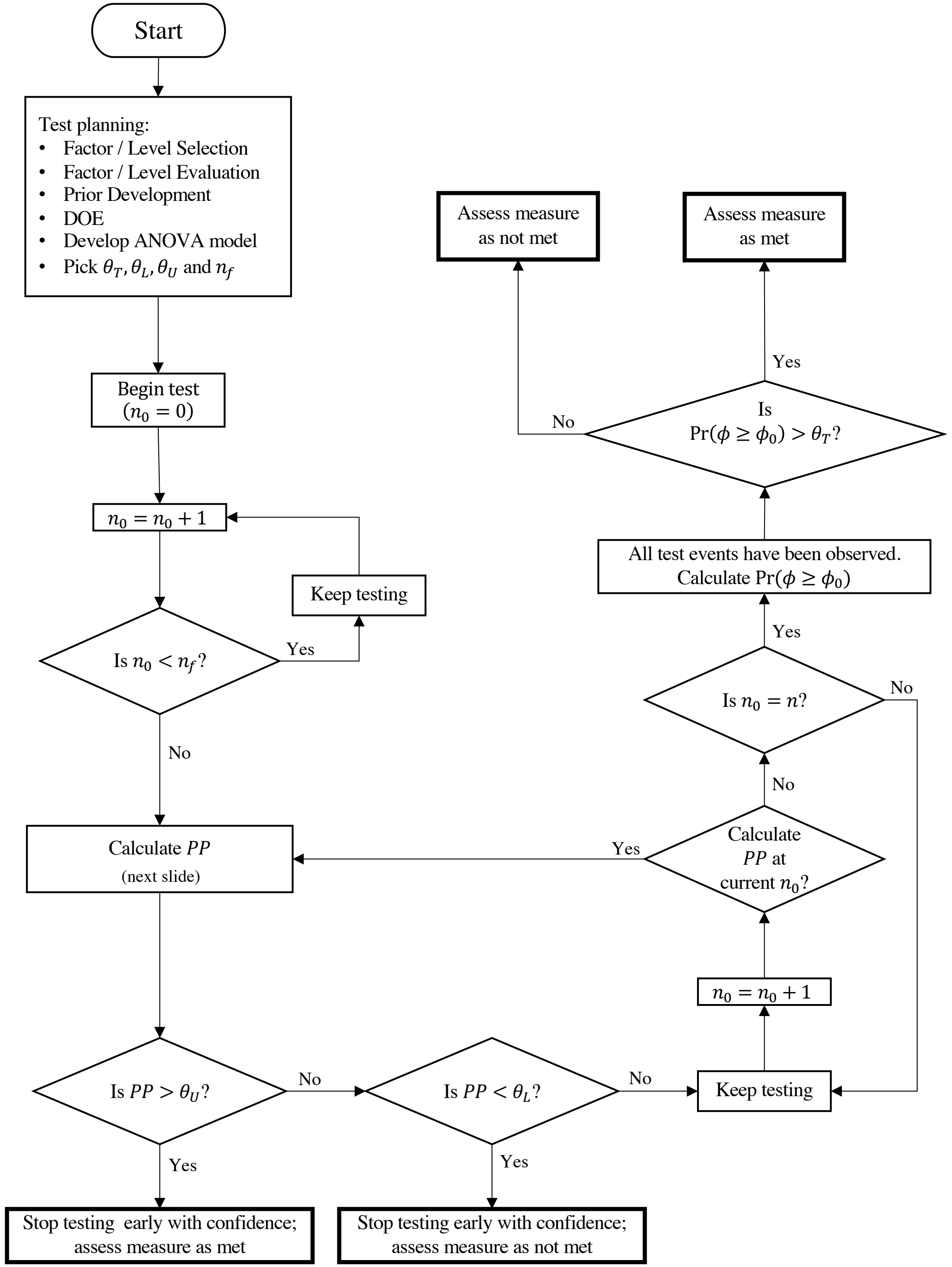
Adaptive Operational Testing: Interim Analysis

Adaptive Operational Testing: Interim Analysis

- If we accomplished the entire experimental design for the test
  - Evaluate a measure as met if $\Pr_{\phi|S}(\phi \geq \phi_0) > \theta_T$, for all seen data, S, and for some threshold value, θ_T
  - Evaluate a measure as not met if $\Pr_{\phi|S}(\phi \geq \phi_0) \leq \theta_T$
- In a Bayesian framework, inferences are constantly updated as data are obtained
- We can determine if a test could end early based on ϕ and test hypothesis, using the predictive probability of evaluating a measure as met at test completion (PP)
  - Lee and Liu (2008) PP proposed for a binomial data model / beta prior
  - Liu and Dressler (2018) extended it to a continuous response with a recognizable posterior
  - Zhou et al. (2018) suggested general framework for using PP for such a case, but did not use it
- Two examples of clinical trials that successfully incorporated PP: I-SPY 2 (on-going) and a completed drug trial adding trastuzumab to chemotherapy

Adaptive Operational Testing

- Using the Bayesian framework, what if we can see interim data?
- Introduce θ_L, θ_U, and PP into the analysis
- Establish rules for when PP can be calculated
  - Frequency of interim data
  - Number of observations to see before calculating PP (n_f)

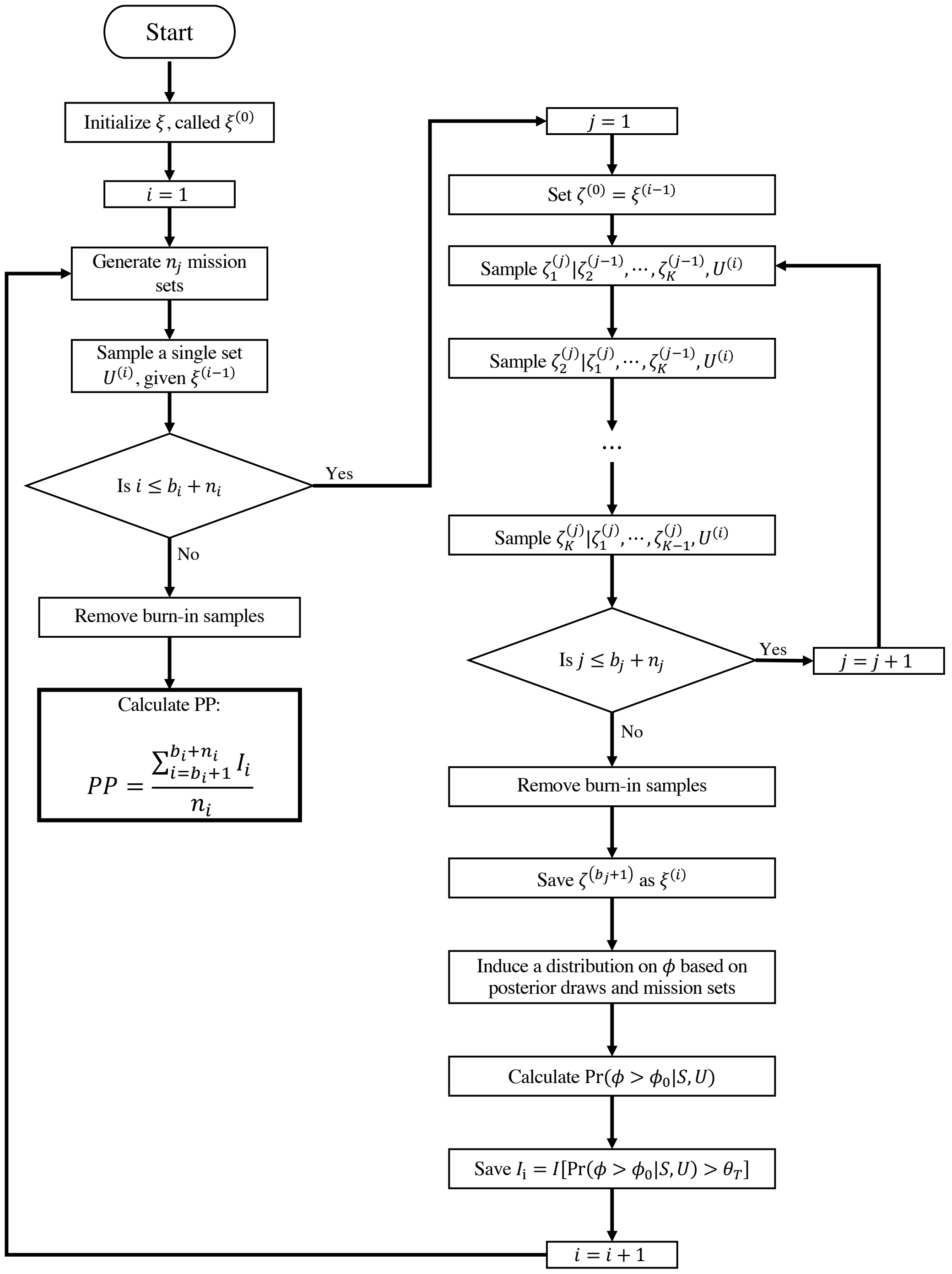


Adaptive Operational Testing: Calculating PP

- PP: For any already seen data, S, and any unseen data, U, what is $\Pr(\phi > \phi_0|S, U)$, and does it exceed θ_T?
- PP is calculated by marginalizing over all possible values of U

$$PP = \Pr(Y\colon \Pr(\phi \geq 400|S, U) > \theta_T)$$
$$= E\{I[\Pr(\phi \geq 400 \mid S, U) \geq \theta_T]|S\}$$

- OT requires a more complex sampling method than currently being implemented



Results Using Weakly Informative Priors

Electric Semi-Truck Example: Proposed Analysis Set-Up

Factor	Levels	Model Parameter
Terrain	Hilly Flat	α_i
Temperature	Hot (> 85º F) Moderate (85º - 50º F)	β_j
Wind	Good Moderate Poor	ω_k
Payload Type	Refrigerated Non-Refrigerated	γ_l
Payload Weight	Heavy (≥ 40k lbs) Light(< 40k lbs)	δ_m

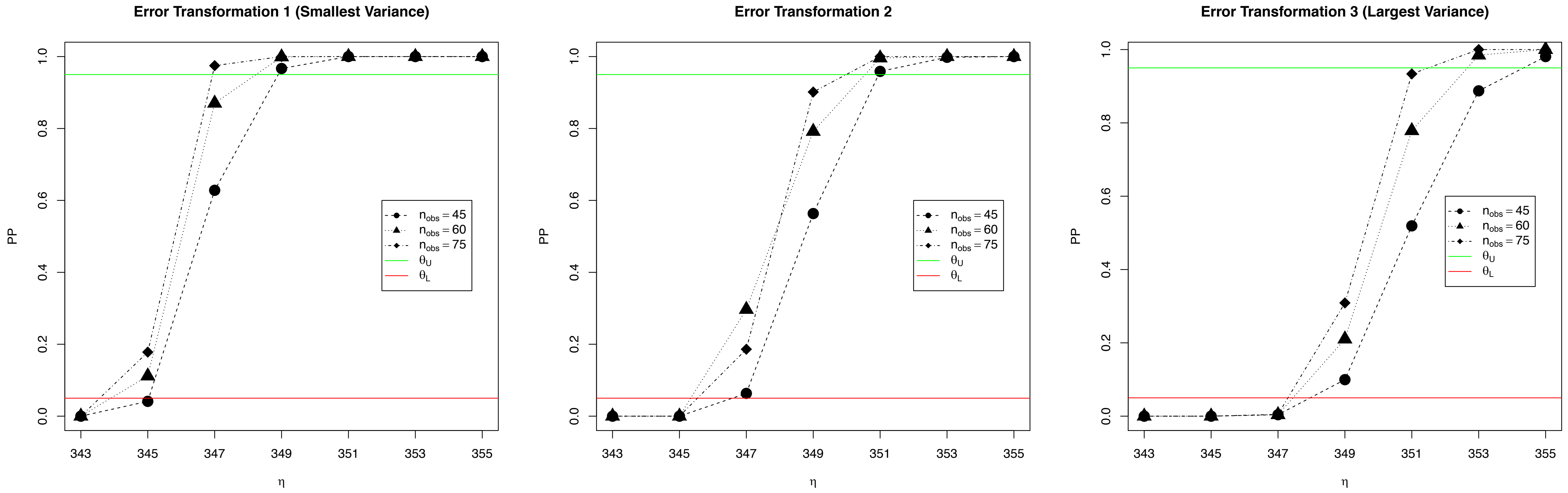
- 2^4 full factorial, five replicates = 80 test events
- Distribution of miles traveled within groups:

$$y_{ijklmp} | \mu_{ijklm} \sim N\left(\mu_{ijklm}, \frac{1}{\tau}\right), \text{ where}$$

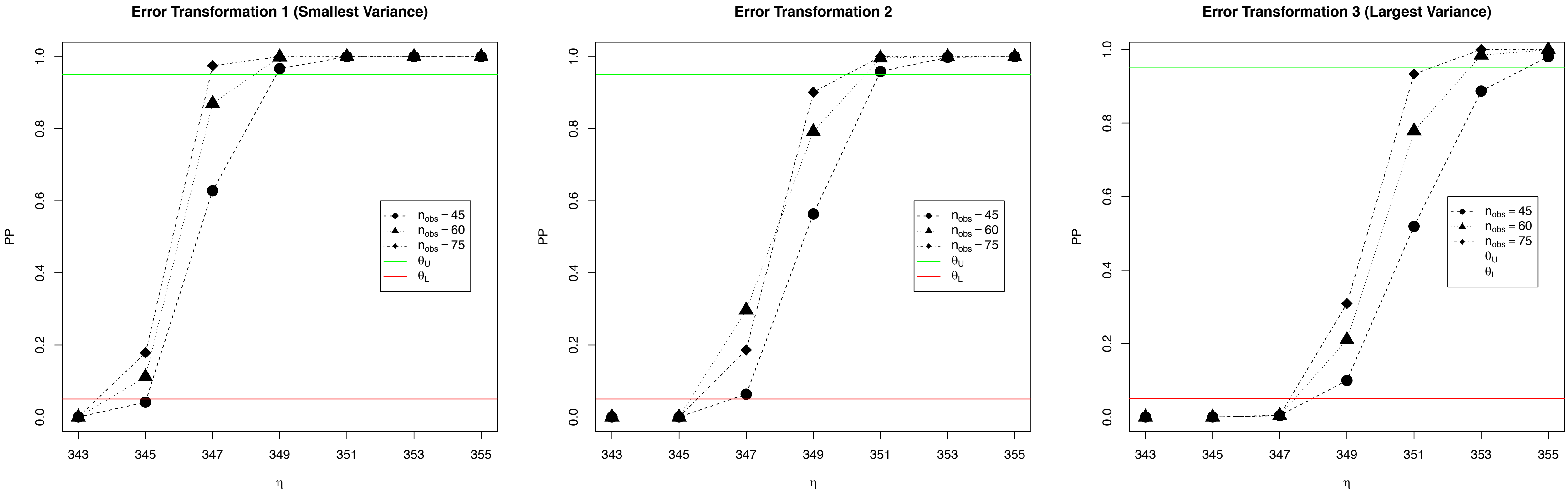
$$\mu_{ijklm} = \eta + \alpha_i + \beta_j + \omega_k + \gamma_l + \delta_m + (\alpha\beta)_{ij} + (\alpha\gamma)_{il} + (\alpha\delta)_{im} + (\beta\gamma)_{jl} + (\beta\delta)_{jm} + (\gamma\delta)_{lm}$$

- OT Data Sets: 21 data sets, changing…
  - true η value (seven different values)
  - the $N(0,1)$ errors (three different transformations)
- Data examples
  - A baseline group (η) observation in Data set 3 comes from a $N(347, 50^2)$
  - A flat (otherwise baseline) group observation in Data Set 3 comes from a $N(397, 54^2)$
  - A baseline group observation in Data set 10 comes from a $N(347, 100^2)$
  - A flat (otherwise baseline) group observation in Data set 10 comes from a $N(397, 109^2)$

Weakly Informative Prior (WIP) Results



Weakly Informative Prior (WIP) Results



Compared to using posterior probability after test completion, the proposed process using *PP* correctly ends testing early in 17 of 21 cases

Key Takeaway: outside a narrow range of η, PP is conclusive

Informative Priors

Creating Informative Priors by Incorporating Previous Information

- Instead of weakly informative priors, informative priors can be created using previous test data, making use of available information

- Using subject matter expert opinion to build informative priors (See Bedrick, Christensen, and Johnson, 1996)
- Summary statistics from previous (related, but dissimilar) tests (See Dewald, Holcomb, Parry, and Wilson, 2016)
- Traditional Approach – exchangeable data
- Variants of power priors – related, but non-exchangeable data

Power Prior Variants

- Normalized Power Prior (NPP):
  - Data determines how much the previous data is down-weighted to account for dissimilarities
  - Model parameters have to be the same in both the historical and current data models

$$p(\boldsymbol{\theta}, a_0 | D_0) \propto \left[\frac{\left(L(\boldsymbol{\theta}|D_0)\right)^{a_0} \pi_0(\boldsymbol{\theta})}{\int \left(L(\boldsymbol{\theta}|D_0)\right)^{a_0} \pi_0(\boldsymbol{\theta}) d\boldsymbol{\theta}} \right] \pi_0(a_0) I_A(a_0)$$

  - See Duan, Ye, and Smith (2006)

- Normalized Partial Borrowing Power Prior (NPBPP):
  - Relaxes the NPP assumption that the model parameters have to be the same in both the historical and current data models
  - Requires a more computationally inefficient Metropolis-within-Gibbs sampler
  - See Chen, Ibrahim, Lam, Yu, and Zhang (2011)

- Conditional Normalized Partial Borrowing Power Prior (CPBPP):
  - Proposed variant of NPBPP that is more computationally efficient
  - Manuscript in preparation

Conclusion

- Ultimately, moving into a Bayesian framework provides OT more flexibility by allowing testers to implement interim analysis.
- Interim analysis allows testers to stop testing early when enough information has been obtained, saving both time and resources (cost).
- The method can be adjusted to incorporate information from previous testing

Questions?

vcarrillo314@unm.edu