FIM Based Power Calculations for NLMEMs

Sebastian Ueckert Andrew C. Hooker

Thursday, September 11, 2014

Power & design in non-linear mixed effect models

- Importance of design for precise parameter estimates should be well known
- In late phase trials statistical power maybe more important
- For non-linear mixed effect models (NLMEMs), relationship between parameter precision and statistical power maybe less obvious

Objectives

- Derive different formulas for calculating power in NLMEMs
- Demonstrate practical application (R code)
- Evaluate the performance of methods

Non-linear mixed effect models

$$y_{ij} = f(g(\theta, \eta_i, a_i), t_{ij}) + h(g(\theta, \eta_i, a_i), t_{ij}, \varepsilon_{ij})$$

y_{ij} Observations j for subject i

θ Fixed effects

η_i Subject specific random effects $(N(0, \Omega))$

a_i Covariates

t_{ij} Observation times

ε_{ij} Random variable describing the residual error $(N(0, \Sigma))$

Running Example - Alzheimer's Disease Trial

- Evaluate novel treatment for Alzheimer's disease (AD)
- Promising results in preclinic
- Want to perform phase IIb POC study
- AD disease progression model from literature

AD Trial - NLMEM

- Disease progression model for ADAS-cog score of subject i at time j:

$$y_{ij} = S_{0_i} + \alpha_i(1 - \gamma \cdot a_{g_i})t_{ij} + A\frac{k_r}{k_r - k_o}(e^{-k_o t_{ij}} - e^{-k_r t_{ij}}) + \varepsilon_{ij}$$

  - Linear natural history
  - Placebo effect according to inverse Bateman function
  - Disease modifying drug effect (30%)
  - Treatment group indicator a_{g_i}

- Between subject model:

$$S_{0_i} = S_0 + \eta_{1i} \qquad \alpha_i = \alpha + \eta_{2i}$$

- Fixed effect parameters

$$\theta = (S_0, \alpha, \gamma, A, k_o, k_r)$$

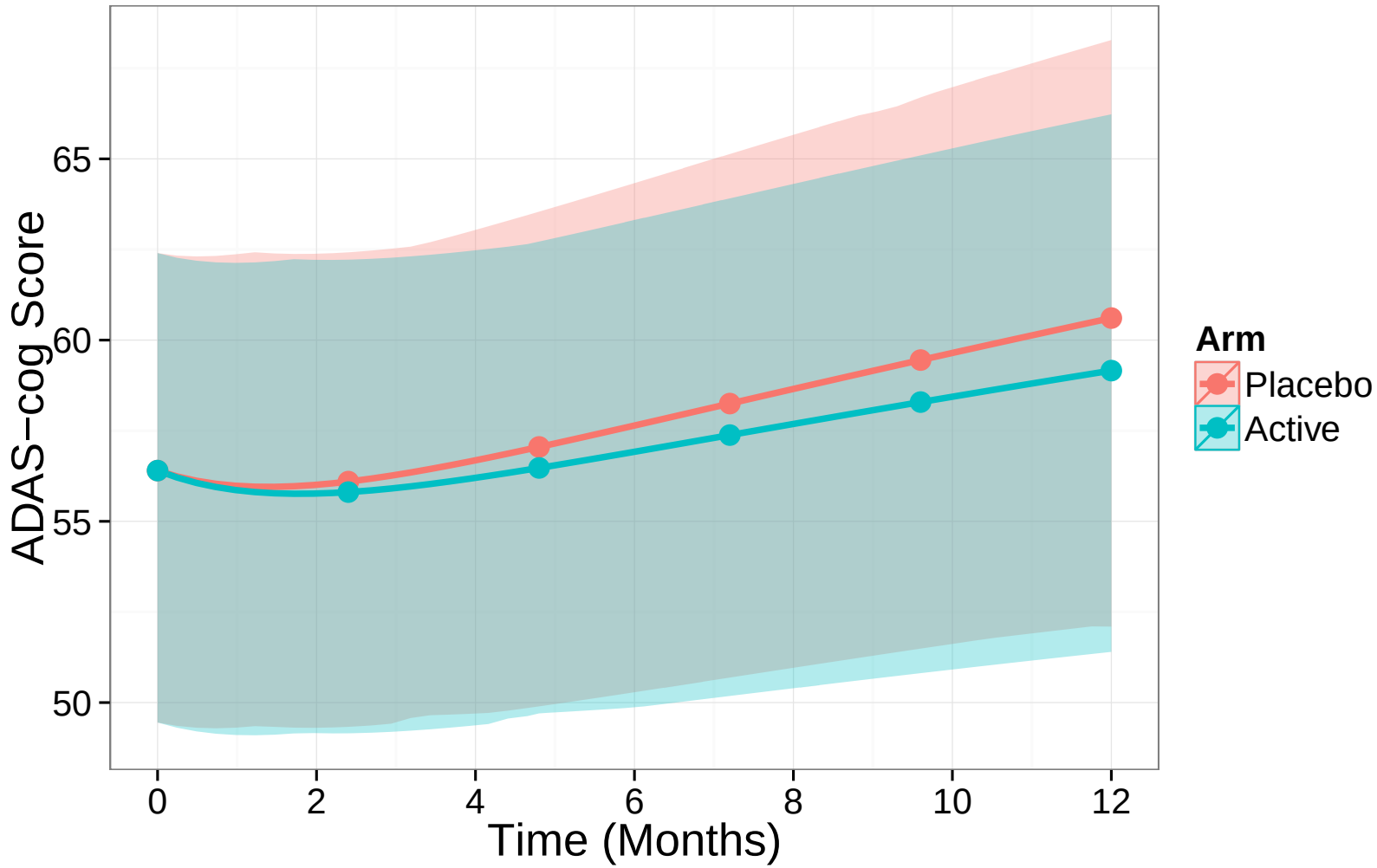
AD Trial - Design

Proposed design:

- 12 months trial
- 2 arms (placebo & active)
- 50 subjects per arm
- 6 observations per subject (equally spaced)

AD Trial - Simulation

Median and 95% prediction interval



Hypotheses tests

Elegant way of making decisions, e.g.

- Structural model building
- Covariate inclusion
- **Drug effect testing**

Often formulated in terms of parameters

$$H_0 : \theta_E = \theta_E^0 \quad H_1 : \theta_E \neq \theta_E^0$$

Define:

- Reduced model f_r: model with $\theta_E = \theta_E^0$
- Full model f_f: model with unrestricted parameters

AD trial - Hypotheses

$$H_0 : \gamma = 0 \qquad H_1 : \gamma \neq 0$$

Full model:

$$y_{ij} = S_{0_i} + \alpha_i(1 - \gamma \cdot a_{g_i})t_{ij} + A\frac{k_r}{k_r - k_o}(e^{-k_o t_{ij}} - e^{-k_r t_{ij}}) + \varepsilon_{ij}$$

Reduced model:

$$y_{ij} = S_{0_i} + \alpha_i t_{ij} + A\frac{k_r}{k_r - k_o}(e^{-k_o t_{ij}} - e^{-k_r t_{ij}}) + \varepsilon_{ij}$$

Wilk's likelihood ratio statistic

Common metric to evaluate evidence against H_0:

$$2 \log \frac{L(\hat{\theta}_E, y)}{L(\hat{\theta}_E^0, y)} = W$$

Generally, require considerable evidence against H_0 (small type 1 error):

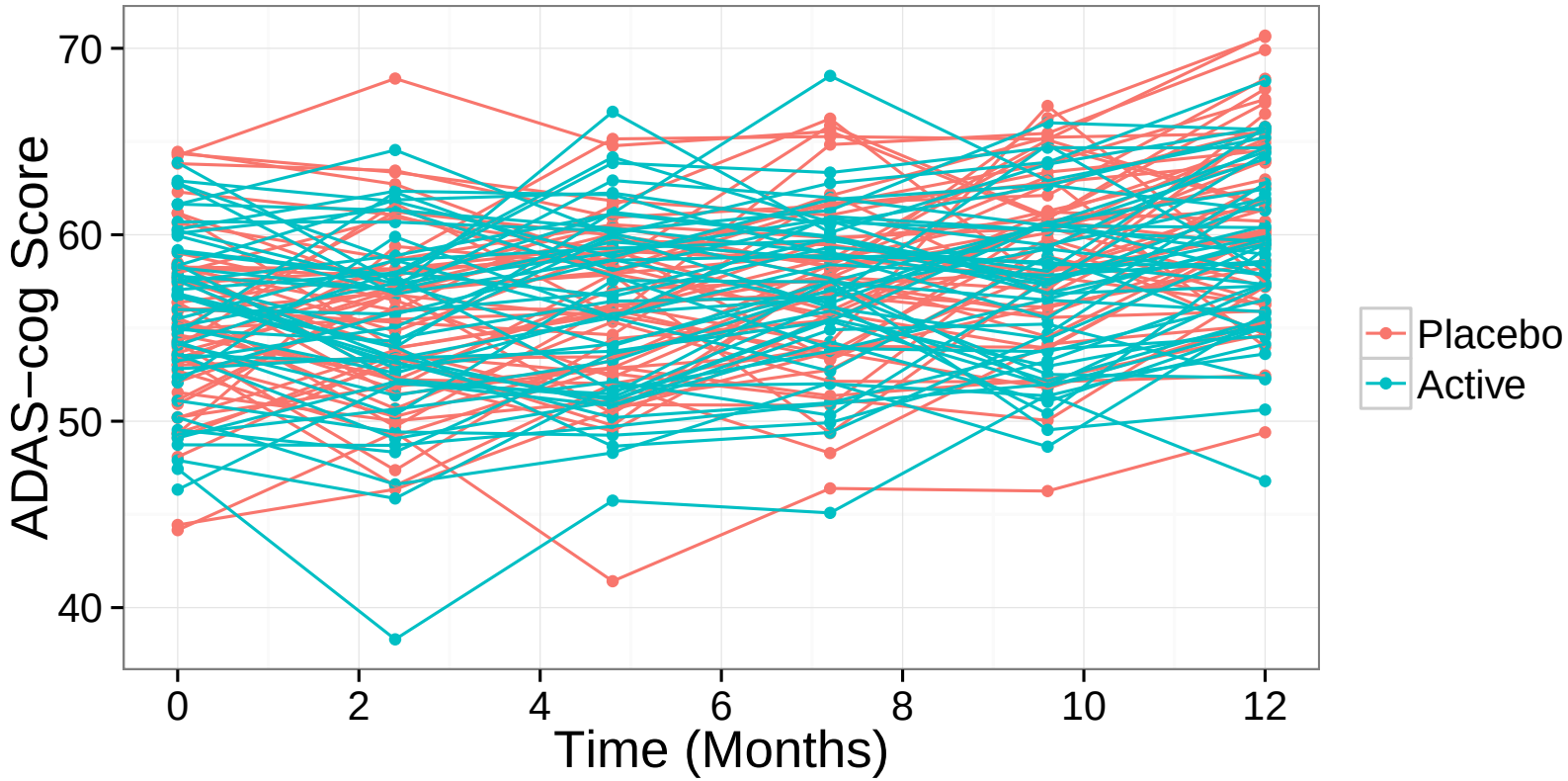
$$\Pr\left(W \geq c | H_0\right) \leq \alpha$$

Often use:

$$\Pr\left(W = w | H_0\right) \xrightarrow{d} f_{\chi^2}(k, w)$$

Take c as quantile from chi-square distribution

AD trial - Testing for a drug effect



$$2\log\frac{L(\hat{\theta}_E, y)}{L(\hat{\theta}_E^0, y)} = 2.54 < 3.84 \rightarrow \text{Not significant}$$

(evaluated using FOCE in NONMEM)

Power

Citing Fisher

> *To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.*

- H_1 generally believed to be true
- Maybe even guess (θ_E) about true value

Before doing experiment, consider

$$\Pr\left(W \geq \chi^2_{1,(1-\alpha)} | H_1\right) = \pi$$

Calculating Power for NLMEMs

$$\Pr\left(W \geq \chi^2_{1,(1-\alpha)} | H_1\right) =$$

$$\Pr\left(2 \log \frac{L(\hat{\theta}_E, y)}{L(\hat{\theta}^0_E, y)} \geq \chi^2_{1,(1-\alpha)} | H_1\right) =$$

$$1 - F_W(\chi^2_{1,(1-\alpha)})$$

$F_W(\chi^2_{1,(1-\alpha)})$ seems complicated and dependent on the data

$\rightarrow$ Monte Carlo simulations

Monte Carlo based power

1. Use Monte-Carlo simulations to generate $y_1, \ldots, y_M$
2. Estimate all y_i with full and reduced model $\rightarrow w_i$
3. Approximate $F_W(x)$ through the empirical distribution function, i.e.

$$F_W(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{w_i \leq x\}$$

4. Calculate power using $\pi_{MC} = 1 - F_W(\chi^2_{1,(1-\alpha)})$

AD trial - Monte Carlo based power

- Implemented AD model in NLMEM software (NONMEM)
- Simulated 500 datasets
  - Estimated[1] with full model
  - Estimated with reduced model
- Calculated the log-likelihood ratios

```
load("../../data/dofv_mc_500.Rdata")
power.mc <- mean(dofv>=qchisq(0.95, 1))
power.mc
```

```
## [1] 0.558
```

[1] FOCE

Summary - Monte Carlo based power

- Intuitive (replicates data analysis process)
- Potentially slow (especially for NLMEMs and power vs. sample size curves)

Alternatives?

Asymptotic distribution under H_1

Under H_1 the W asymptotically follows a non-central chi-square distribution[2],i.e.

$$\Pr\left(W = w | H_1\right) \xrightarrow{d} f_{\tilde{\chi}^2}(\lambda, k, w)$$

$f_{\tilde{\chi}^2}(\lambda, k, w)$ is pdf of non-central chi-square distribution

- Degrees of freedom k
- Non-centrality parameter λ

Know k, but not $\lambda \rightarrow$ estimate it

[2]Rochon (1998)

Parametric power estimation

1. Use Monte-Carlo simulations to generate $y_1, \ldots, y_M$
2. Estimate all y_i with full and reduced model $\rightarrow w_i$
3. Estimate $\hat{\lambda} = \text{argmax} \sum_{i=1}^{M} \log f_{\tilde{\chi}^2}(\lambda, k, w_i)$

$$\pi_{PPE} = 1 - \int_{-\infty}^{\chi^2_{1,(1-\alpha)}} f_{\tilde{\chi}^2}(\hat{\lambda}, k, x) dx$$

AD trial - Parametric power estimation

- Implemented AD model in NLMEM software (NONMEM)
- Simulated 500 datasets
  - Estimated[3] with full model
  - Estimated with reduced model
- Calculated the log-likelihood ratios

```
load("../../data/dofv_mc_500.Rdata")
ll<-function(ncp)-sum(dchisq(dofv,df=1,ncp,log=T))
fit <- optim(par=mean(dofv)-1, fn=ll, method="BFGS")
power.ppe <- 1-pchisq(qchisq(0.95,df=1),df=1,
                      ncp=fit$par)
power.ppe
```

```
## [1] 0.5211
```

[3] FOCE

Summary - Parametric power estimation

- Easy to implement
- Conceptually more complex
- More assumptions
- Still dependent on simulated data

Removing the data dependence

Relationship for the non-centrality parameter[4]

$$\lambda = \Psi(\hat{\theta})^T \left[\frac{\partial \Psi(\hat{\theta})}{\partial \theta} I(\hat{\theta})^{-1} \frac{\partial \Psi(\hat{\theta})^T}{\partial \theta} \right]^{-1} \Psi(\hat{\theta})$$

- Ψ constraint function such that $\Psi(\theta) = 0$ under H_0
- $I(\hat{\theta})$ observed information matrix

Substituting I with the expected information matrix $\mathcal{I}$ and $\hat{\theta}$ with θ^G

$$\lambda \approx \Psi(\theta^G)^T \left[\frac{\partial \Psi(\theta^G)}{\partial \theta^G} \mathcal{I}(\theta^G)^{-1} \frac{\partial \Psi(\theta^G)^T}{\partial \theta^G} \right]^{-1} \Psi(\theta^G)$$

[4] Rochon (1998)

Wald statistic

For hypotheses of the form

$$H_0 : \theta_E = \theta_E^0 \quad H_1 : \theta_E \neq \theta_E^0$$

and effect guess θ_E^G

$$\lambda_W = \Psi(\theta_E^G)^T \left[\frac{\partial \Psi(\theta_E^G)}{\partial \theta_E^G} \mathcal{I}(\theta_E^G)^{-1} \frac{\partial \Psi(\theta_E^G)^T}{\partial \theta_E^G} \right]^{-1} \Psi(\theta_E^G) =$$

$$(\theta_E^G - \theta_E^0) \mathcal{I}_(\theta_E^G)^{-1} (\theta_E^G - \theta_E^0)^T$$

Wald based power

1. Calculate the expected information matrix $\mathcal{I}$
2. Calculate non-centrality parameter λ_W using

$$\lambda_W = (\theta_E^G - \theta_E^0)\mathcal{I}_(\theta_E^G)^{-1}(\theta_E^G - \theta_E^0)^T$$

3. Calculate power using the cdf of non-central chi square distribution, i.e.

$$\pi_W = 1 - \int_{-\infty}^{\chi^2_{1,(1-\alpha)}} f_{\tilde{\chi}^2}(\lambda_W, k, x)dx$$

Wald based power curve

Information matrix $\mathcal{I}$ is additive for different subjects, i.e.

$$\mathcal{I}(\Xi) = N\mathcal{I}(\xi_0)$$

if design is the same for all subjects N

$$\lambda_W(n) = n(\theta_E^G - \theta_E^0)\mathcal{I}_0(\theta_E^G)(\theta_E^G - \theta_E^0)^T$$

$\rightarrow$ Directly obtain power versus sample size curve

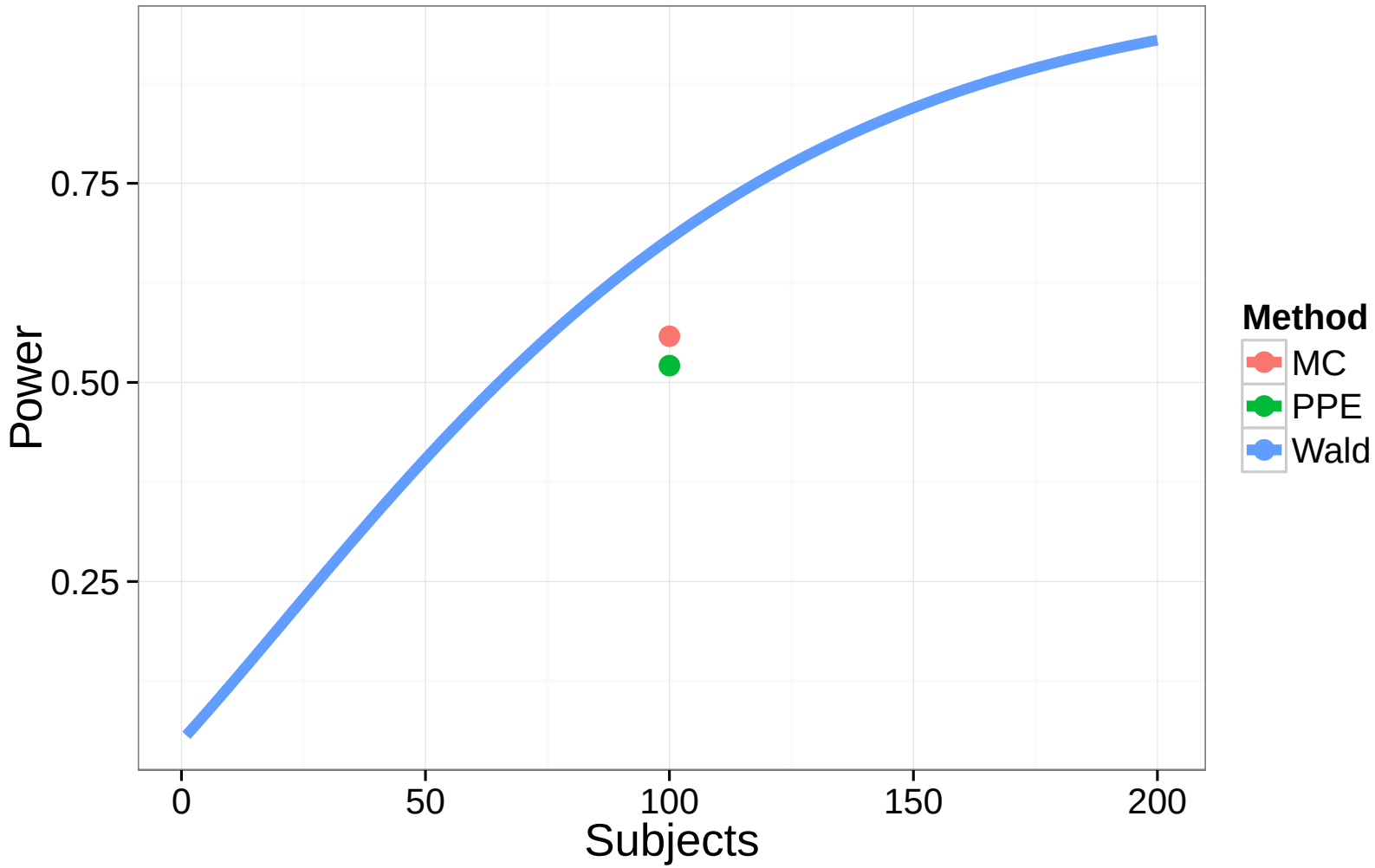
AD trial - Wald based power

- Implemented AD model in PopED for R

```
source("../../scripts/ad_example.R")
fim <- evaluate.fim(poped.db)
ncp <- get_all_params(poped.db)$bpop[3]^2/solve(fim)[3,3]
power.w <- function(n) 1-pchisq(qchisq(0.95,df=1),
                                 df=1,ncp=ncp*n/100)
power.w(100)
```

```
## [1] 0.6802
```

AD trial - Wald based power curve



Summary - Wald based power

- Very fast
- Delivers full power curves
- Requires information matrix (discrete NLMEMs?)
- Optimistic

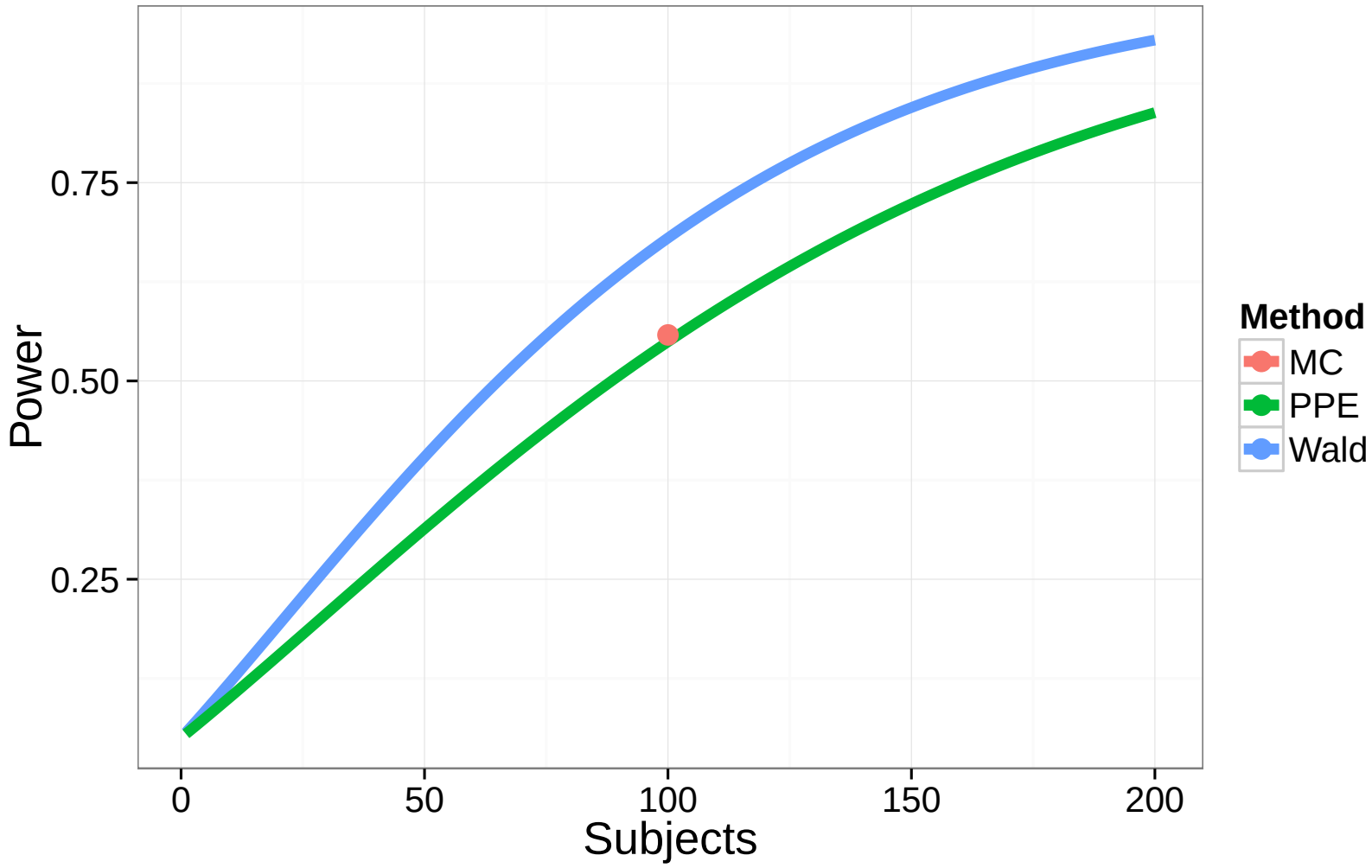
PPE based power curve

Relationship between study size and non-centrality parameter can also be used for parametric power estimation

1. Use Monte-Carlo simulations to generate $y_1, \ldots, y_M$
2. Estimate all y_i with full and reduced model $\rightarrow w_i$
3. Estimate $\hat{\lambda}_0 = \text{argmax} \sum_{i=1}^{M} \log f_{\tilde{\chi}^2}(\lambda, k, w_i)$
4. For every sample size n calculate power using

$$\pi_{PPE}(n) = 1 - \int_{-\infty}^{\chi^2_{1,(1-\alpha)}} f_{\tilde{\chi}^2}(\hat{\lambda}_0 \frac{n}{N_0}, k, x)dx$$

AD trial - PPE based power curve



Non-linear Wald statistic

Wald statistic only considers precision of effect parameters (parameters constrained under H_0)

$$(\theta_E^G - \theta_E^0)\mathcal{I}_(\theta_E^G)^{-1}(\theta_E^G - \theta_E^0)^T$$

But, likelihood ratio statistic takes all parameters into account

Consider instead

$$H_0 : E(f_f) = E(f_r) \qquad H_1 : E(f_f) \neq E(f_r)$$

i.e. under H_0 full and reduced model predictions are identical (non-linear hypothesis)

Non-linear Wald statistic (2)

Remember formula for λ

$$\lambda \approx \Psi(\theta)^T \left[\frac{\partial \Psi(\theta)}{\partial \theta} \mathcal{I}(\theta)^{-1} \frac{\partial \Psi(\theta)^T}{\partial \theta} \right]^{-1} \Psi(\theta)$$

Derive Ψ using first order approximation

$$E(f(g(\theta, \eta_i, a_i), t_{ij})) \approx f(g(\theta, \eta_i, a_i), t_{ij})|_{\eta_i=0}$$

$$\Psi(\theta) = (f_f - f_r)\,|_{\eta_i=0} = 0$$

Combining with λ formula

$$\lambda_{NLW} = (f_f - f_r)^T \left[\frac{\partial (f_f - f_r)}{\partial \theta} \mathcal{I}(\theta)^{-1} \frac{\partial (f_f - f_r)}{\partial \theta} \right]^{+} (f_f - f_r)$$

(+ Moore-Penrose generalized inverse)

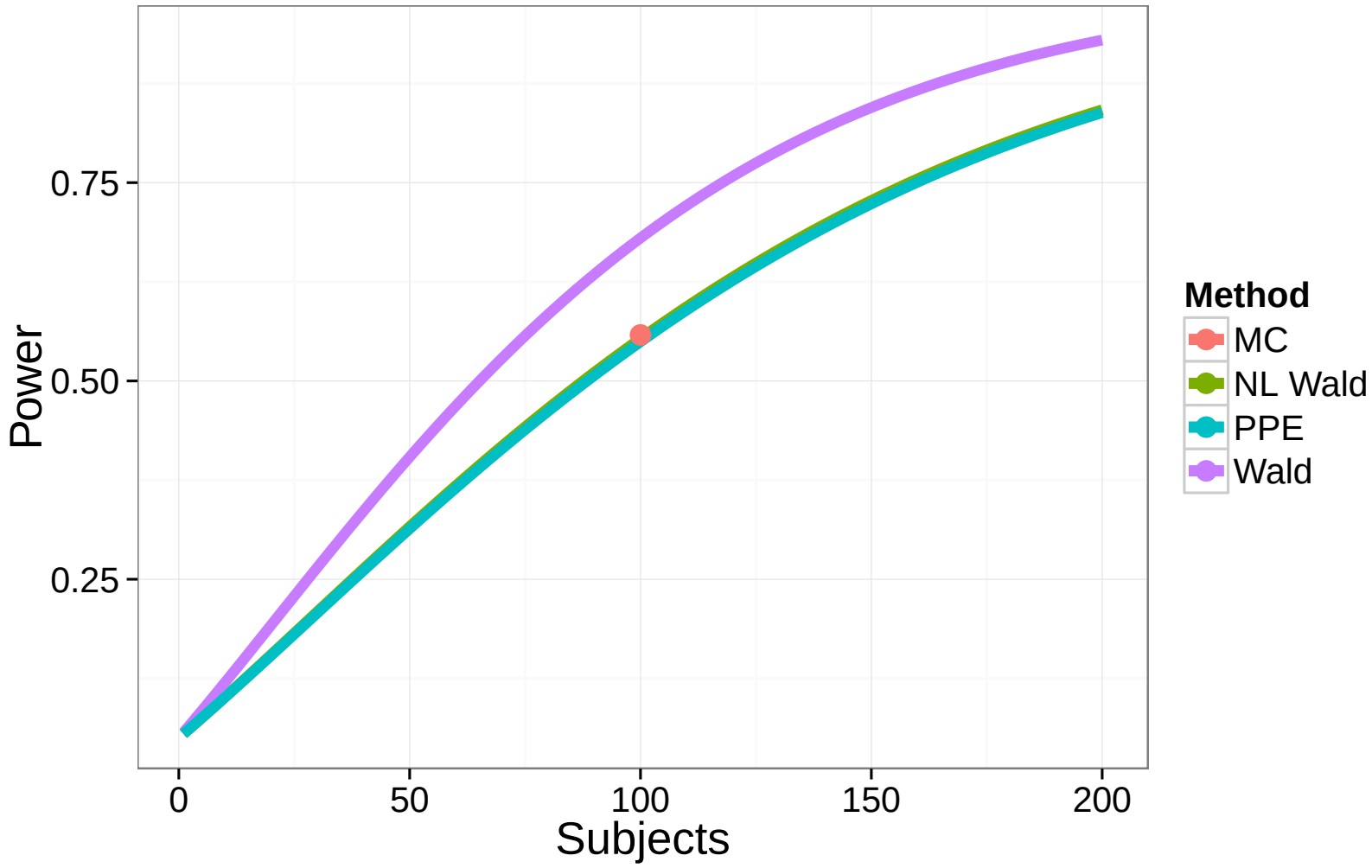
AD trial - Non-linear Wald power

- Implemented AD model in PopED for R
- Calculated f_f, f_r, $\partial f_f/\partial\theta$ and $\partial f_r/\partial\theta$

```
fim <- evaluate.fim(poped.db)
dpsi <- dF-dR
psi <- f-r
ncp.nwald <- t(psi)%*%pinv(dpsi%*%
             solve(fim)[fe_index,fe_index]%*%
             t(dpsi))%*%psi
power.nlw <- function(n) 1-pchisq(qchisq(0.95,df=1),
                                  df=1,ncp=ncp.nwald*n/100)
power.nlw(100)
```

```
## [1] 0.5535
```

AD trial - Non-linear Wald power curve



Summary - Non-linear Wald power

- Very fast
- Delivers full power curves
- More precise than "classical" Wald[5]
- Requires information matrix (discrete NLMEMs?)

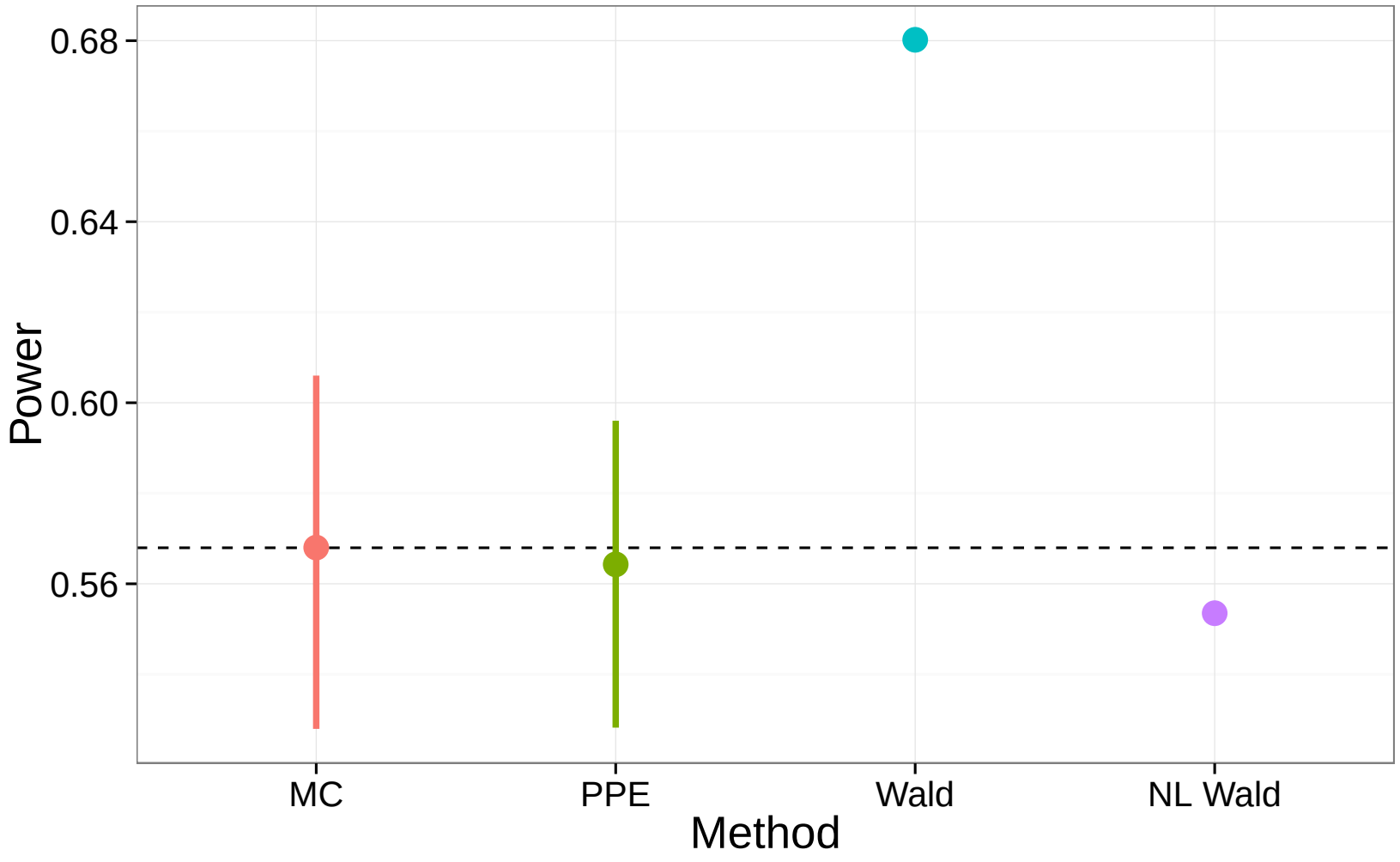
[5] Ueckert et al. 2012

Method comparison

Compare for AD trial scenario:

- MC power based on 10,000 samples (reference)
- MC power based on 500 samples (1000 repetitions)
- PPE power based 500 samples (1000 repetitions)
- Wald power
- NL Wald power

Method comparison - Results



Summary

- Monte Carlo
  - Easy to implement
  - Computationally expensive
- Parametric power estimation
  - Fewer Monte Carlo samples & full power power curves
  - More assumptions
- Wald statistic (linear)
  - Very fast
  - Requires FIM
  - Very optimistic
- Non-linear Wald statistic
  - Very fast
  - More precise than linear Wald
  - Requires FIM
  - Optimistic

Thanks to

Prof. France Mentré for valuable input

All of you for listening