Bayesian aggregation of average data in hierarchical models

S. Weber[1], A. Gelman[2], B. Carpenter[2], D. Lee[2], F. Y. Bois[3], A. Racine[1]

(1) Novartis, Basel; (2) Columbia Uni., New York; (3) Uni. de Technologie de Compiègne

Bayes Pharma, 20. May 2015, Basel, Switzerland

NOVARTIS

Outline

1. Motivation: Trial Design with Hierarchical Models Using Patient Data **and** Data Summaries
2. Bayesian Aggregation of Summary Data
3. Example: Non-Linear Hierarchical Model

NOVARTIS

Designing a Trial with an Active Control Arm

Assessing a Trial Design with Different Sources of Information

- Examples include non-inferiority and bio-similarity trials
- Test of a candidate substance against an active control
  - Candidate substance developed «in-house»
    → lots of raw data - individual patient level longitudinal data
  - Active control developed «externally»
    → only data summaries - publications or submission documents
- New trial will be similar to earlier trials of the active control drug, but still to some extent different
  → meta-analysis on data summaries possible
  - Allows for partial pooling
  - Between-trial heterogeniety often difficult to handle
  - **Restricted to reported summary endpoints and designs**

Drug Disease Modeling of Drug Responses

Key Application is Clinical Trial Simulations (CTS) for Study Design

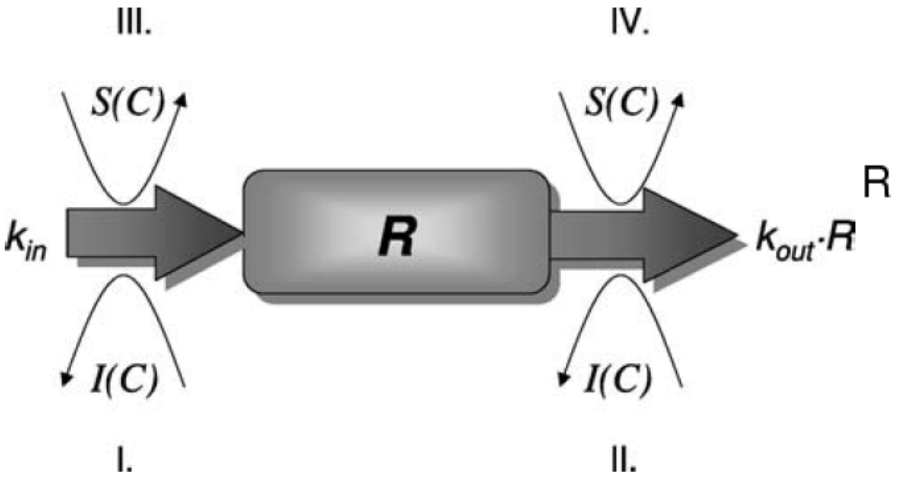
- Simulation of drug responses of patients over time
  - New designs can be considered (regimes/incl.+excl. criteria/...)
  - Different endpoints can be explored including time to event
- Hierarchical (population) based models commonly done →**Requires patient-level data**
- Paradox: Same disease progression, same patient population and likely similar mechanism of action, but population model describes only «in-house» drug

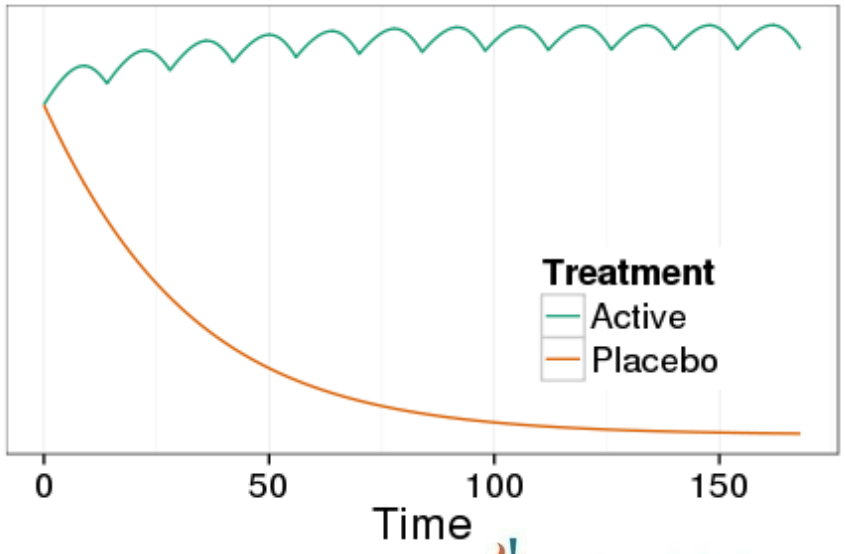
How to learn from published summaries of longitudinal data in the context of non-linear hierarchical models?

Semi-Mechanistic Turn-Over Models

Linking Pharmacokinetics (PK) with Pharmacodynamics (PD)

- PD response *R* can be safety or efficacy related driven by PK effect on «PD bio-compartement»
  - Zero order «production» / first order «elimination» of response R
  - 4 variants: zero / first order inhibition / stimulation due to PK
  - Drug response with respect to reference state (placebo)
- Some regimens may lead to PK causing oscillations and hence oscillations in response





Source: Peletie LA et al.; J Pharmacokinet Pharmacodyn. 2005

NOVARTIS

Example: Simplified «Stimulation of k^{in}»

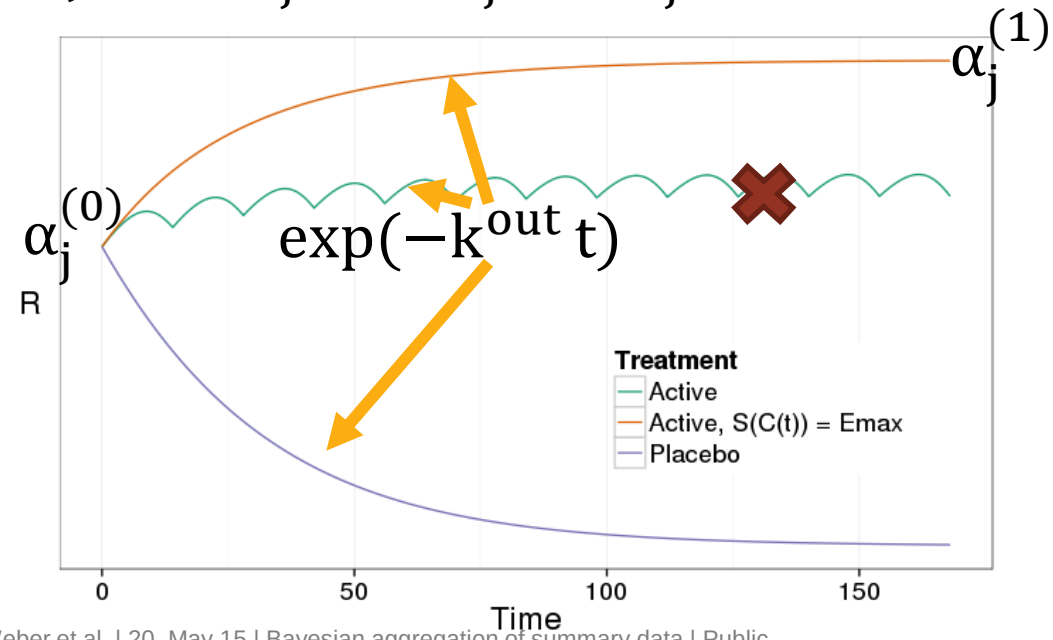
Assuming Maximal Effect at All Times Allows for Analytic Solution

- General turn-over model (only as ODE)

$$\frac{dR_j(t)}{dt} = k^{in}{}_j\,[1 + S_j(C_j(t))] - k^{out}\,R_j(t)$$

- Simplification here: $C_j(t) \gg EC50 \;\Rightarrow\; S_j\big(C_j(t)\big) = E_{max_j}$

$$R_j(t) = \alpha_j^{(1)} + [\alpha_j^{(0)} - \alpha_j^{(1)}]\exp(-k^{out}\,t).$$

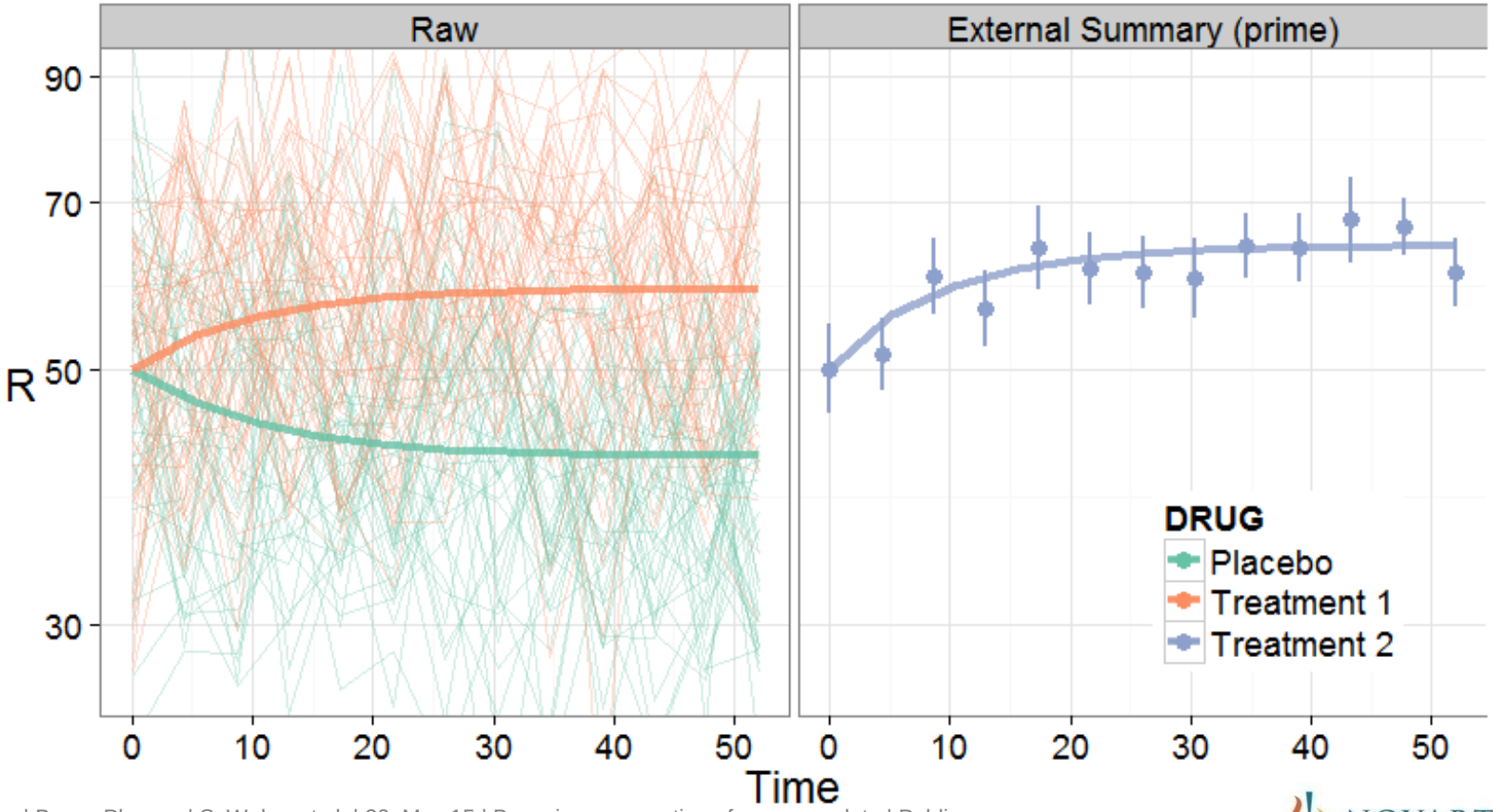


NOVARTIS

Simulated Example Data Set

50 Patients per Treatment Arm Placebo, Treatment 1 & 2

- Solid line is true population mean
- No population differences
- No between-trial variation



NOVARTIS

Integrating External Summary Data

Data Generating Model Must be Shared at Least Partially

- External data $\bar{y}'$ assumed to follow almost the same model
  - Populations must be comparable (or needs appropiate adjustment)
  - Variance components must be identical
  - Natural disease progression must be the same
- Statistical considerations
  - No gain to include external data if ϕ' *completley* unrelated to ϕ
  - If $\delta = \phi' - \phi = 0$ then only precision improvement
  - Information is partially shared, i.e. few components of $\delta \neq 0$
    → Most interesting as scope of model expands
- Direct approach is to consider patients as latent
  → High computational burden (dimensionality increase)
  (see Chiu, A. W. & Bois, F.Y. 2007)

Hierarchical Model Structure

Likelihood on The Basis of Patients; only Population Posterior of Interest

- Notation
  - Observations y from J patients at T time-points

$$y = (y_{jt}; j = 1, \dots, J; t = 1, \dots, T)$$

  - Parameter vector separated in population ϕ and patient α

$$(\phi, \alpha)$$

- Model
  - Prior

$$p(\phi, \alpha) = p(\phi) \prod_{j=1}^{J} p(\alpha_j|\phi)$$

  - Likelihood for patient j

$$p(y_j|\alpha_j, \phi)$$

  - Full posterior

$$p(\phi, \alpha|y) \propto p(\phi) \prod_{j=1}^{J} p(\alpha_j|\phi) \prod_{j=1}^{J} p(y_j|\alpha_j, \phi)$$

Interest is only in the marginal posterior $p(\phi|y)$

Approximation of the Patient Level Likelihood

Key Idea is to Approximate Likelihood using Predictive of the Summary

- Approximation steps to get

$$p(\phi, \delta | y, \bar{y}')$$

  - Full posterior if full raw data would be available

$$p(\phi, \delta, \alpha, \alpha' | y, y') = p(\phi, \alpha | y)\, p(\delta, \alpha' | \phi, y')$$

$$\propto p(\phi, \alpha | y)\, p(\delta | \phi)\, \Pi_{j=1}^{J'} p\,(\alpha'_j | \phi, \delta)\, \Pi_{j=1}^{J'} p\,(y'_j | \alpha'_j, \phi, \delta)$$

  - Approximation of patient level likelihood for the summary

$$\Pi_{j=1}^{J'} p\,(y'_j | \alpha'_j, \phi, \delta) \approx p(\bar{y}' | \phi, \delta)$$

  - Approximated posterior of interest is a **reweighted posterior** from the first inference (just like any Bayesian inference)

$$p(\phi, \delta | y, \bar{y}') \propto p(\phi | y)\, p(\delta | \phi)\, p(\bar{y}' | \phi, \delta)$$

- **Key idea: Replace patient likelihood by probability model of external summary given by the predictive**

Simulation Based Importance Weights

Central Limit Theorem Justifies a Multivariate Normal Approximation

1. Obtain S draws ϕ^s from $p(\phi|y)$
2. For each draw ϕ^s do
   a) Sample from $p(\delta|\phi^s)$; which is then $(\phi^s, \delta^s) \sim p(\phi, \delta|y)$
   b) Simulate $\tilde{J}$ times $\tilde{\alpha}_j$ and data $\tilde{y} = (\tilde{y}_{jt}; j = 1, \ldots, \tilde{J}; t = 1, \ldots, T')$
   Note: $\tilde{J} \neq J'$ and $T \neq T'$ (design of y' maybe different)
   c) Compute mean vector $\tilde{M}^s$ and covariance matrix $\tilde{\Sigma}^s$
   d) Approximate probability model of external summary with a MVN which is the *importance ratio*

$$\boldsymbol{r^s = N(\bar{y}'|\tilde{M}^s, \frac{1}{J'}\tilde{\Sigma}^s)}$$

3. Compute *truncated importance weights* $w^s = \min(r^s, \sqrt{S}\bar{r})$
4. Use importance resampling to obtain final posterior

Efficiency Considerations

Wide Prior on δ Leads to Poor Performance

- Wide δ prior implications
  - Will ensure an overlap with «best-fit» posterior
  - Most simulation draws are wasted
  - «Optimal» δ prior is its posterior in terms of efficiency
- Solution: Use **suitable proposal** g(x) and adjust ratios

$$E_p(x) = \int x\, p(x)\, dx = \int x\, \frac{p(x)}{g(x)}\, g(x)\, dx = E_g(r\, x)$$

  - Sample in 2a) δ^s from density g instead from prior
  - **Adjust** importance ratio for draw s

$$r^s = \frac{p(\delta^s|\phi^s)}{g(\delta^s)}\, N(\bar{y}'|\tilde{M}^s, \frac{1}{J'}\tilde{\Sigma}^s)$$

  - Use **spread out posterior** as new proposal density and **iterate**

Key Assumptions

Shared Model, Normal Summary Probability Model, Efficient Proposal

1. External data generated from **partially shared model**
   In the following example we assume:
   $$\delta = \phi' - \phi = \left(0, \delta^{(2)}, 0, \ldots, 0\right)$$

2. Patient level likelihood for external data replaced by **probability model of external summary**
   $$\Pi_{j=1}^{J'} p\left(y'_j \middle| \alpha'_j, \phi, \delta\right) \approx p\left(\bar{y}' \middle| \phi, \delta\right)$$
   For mean summaries the **central limit theorem** justifies the use of the **multivariate Normal**

3. Efficiency ensured via **suitable proposal density** g
   $$r^s = \frac{p(\delta^s|\phi^s)}{g(\delta^s)} N(\bar{y}'|\tilde{M}^s, \frac{1}{J'}\tilde{\Sigma}^s)$$

Evaluation Strategy

Compare Approximation Raw/Summary vs All Raw Scenario

- Simulation data set
  1. Simulate 50 patient profiles per arm (placebo, treatment 1 & 2)

$$\log(\mathrm{E}_{\max_j}) = \begin{cases} -\infty & j \in \text{placebo} \\ \phi^{(2)} & j \in \text{treatment 1} \\ \phi^{(2)} + \delta^{(2)} & j \in \text{treatment 2} \end{cases}$$

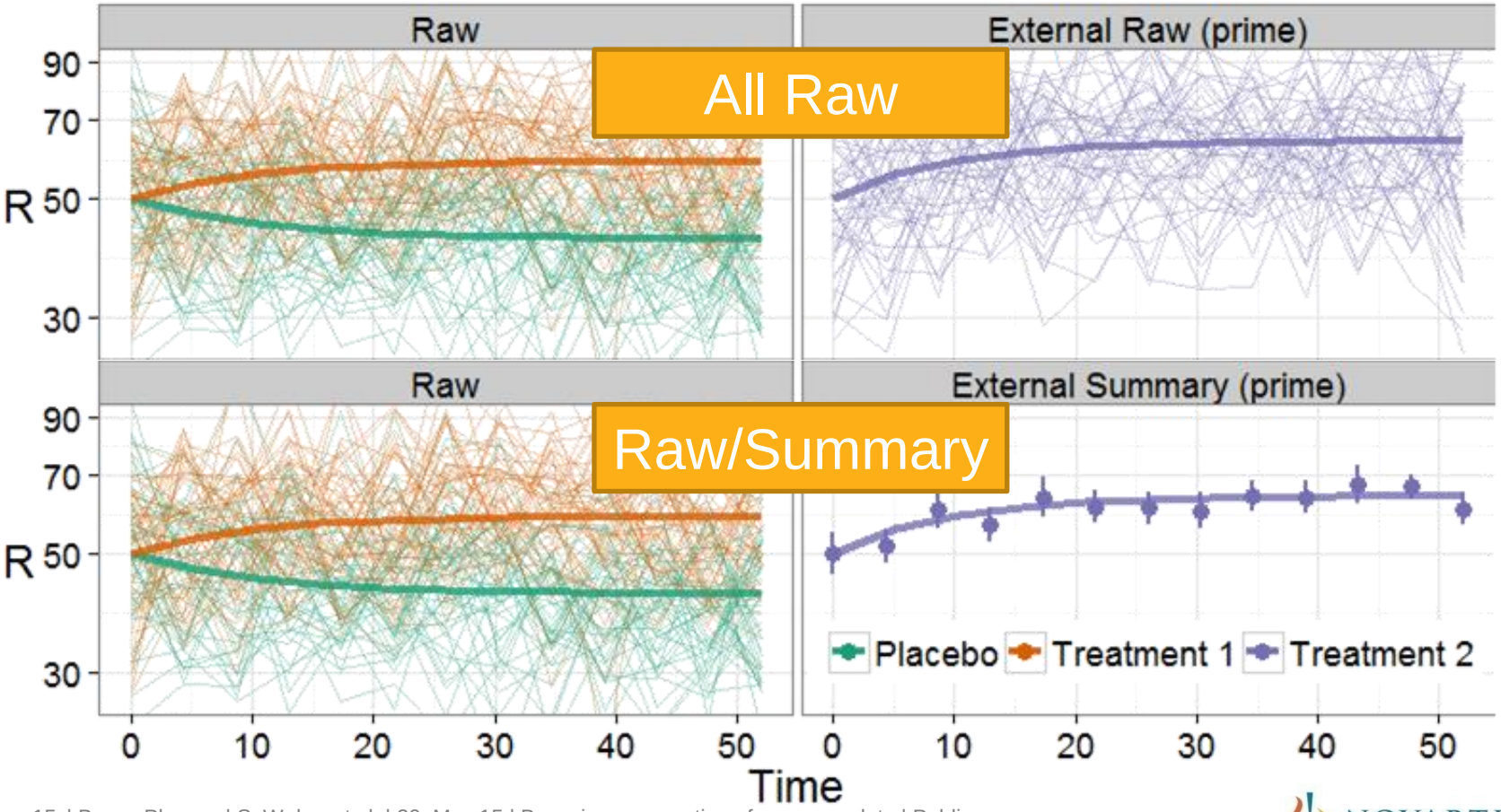
  2. Summarize data for treatment 2 with geometric means
  3. Repeat 1 & 2 for $\delta^{(2)} \in \{0.00, 0.27, 0.55, 0.82, 1.10\}$

- Prior $\delta^{(2)} \sim \mathrm{Normal}(0,10)$
- Evaluation with multiple runs
  - **All Raw**: Full data set on placebo, treatment 1 & 2
  - **Raw/Summary**: Placebo, treatment 1 & summary of treatment 2

Simulation Example

Compare Approximation Raw/Summary vs All Raw Scenario

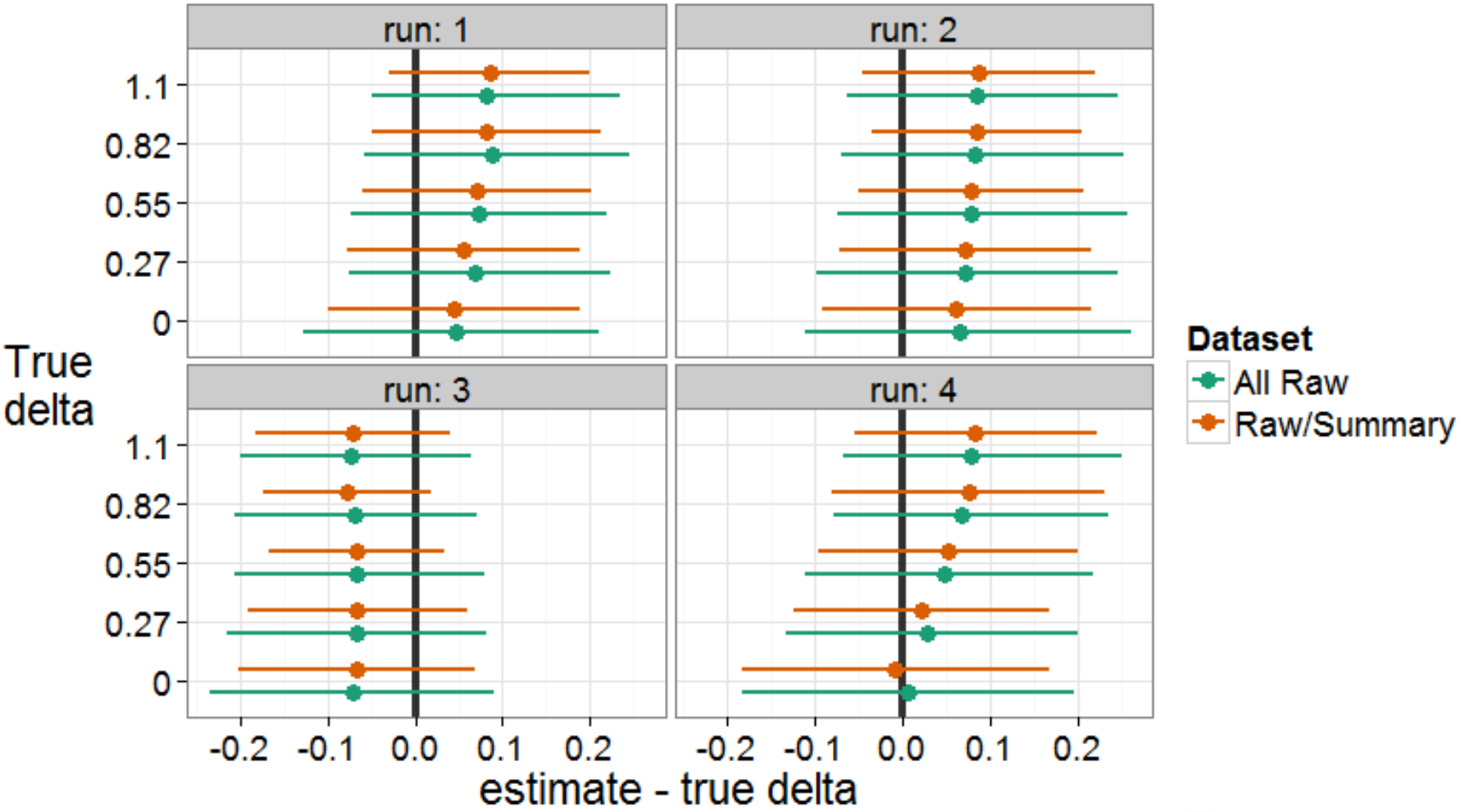
- 50 patients per arm
- 5 true different δ
- External with different Emax
- Multiple runs



NOVARTIS

Evaluating the Estimation of δ

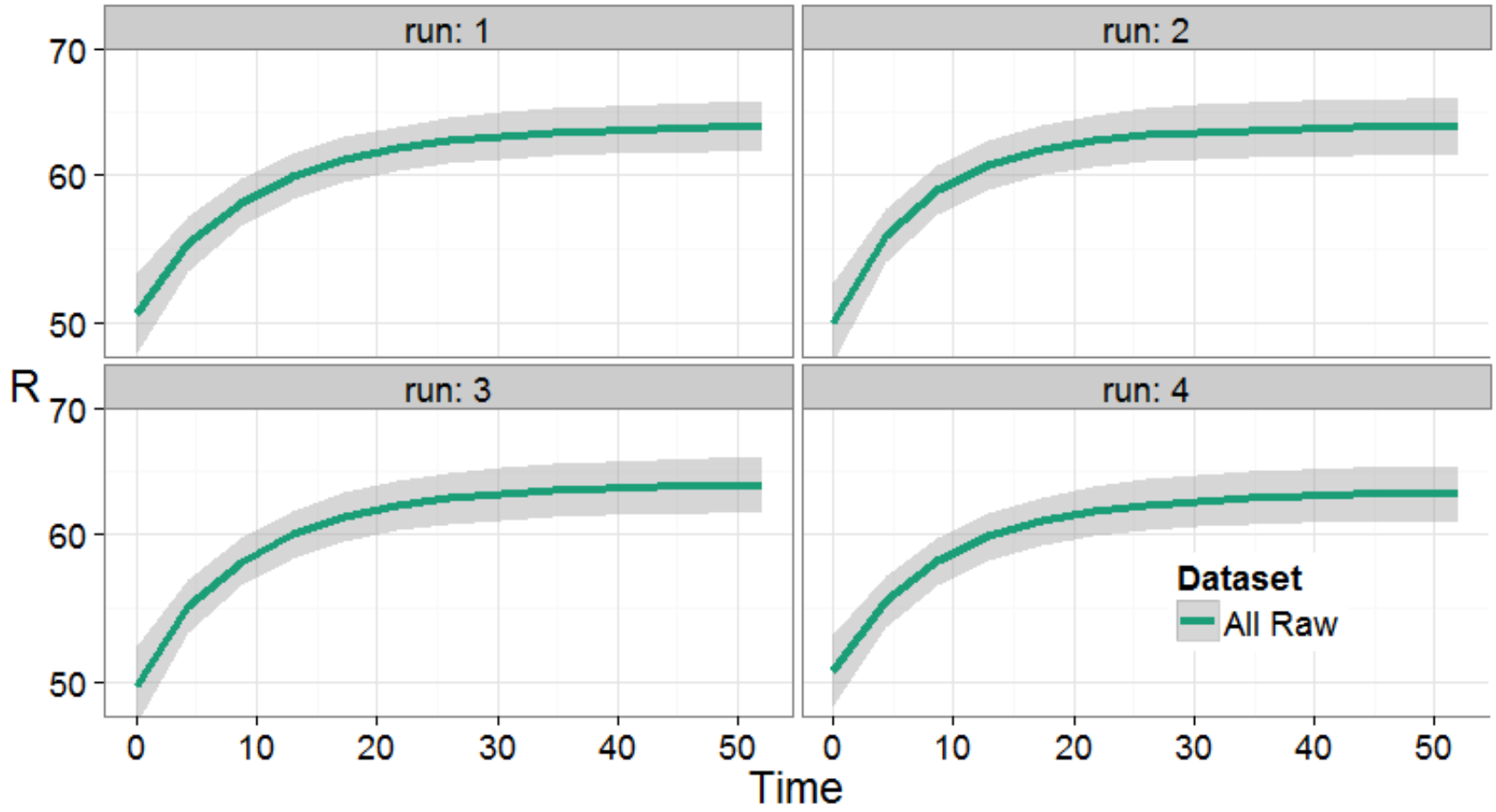
Good Consistency, Coverage Slightly too Small based on 4 Simulations



NOVARTIS

Comparison for the Prediction of the 95% CI Mean

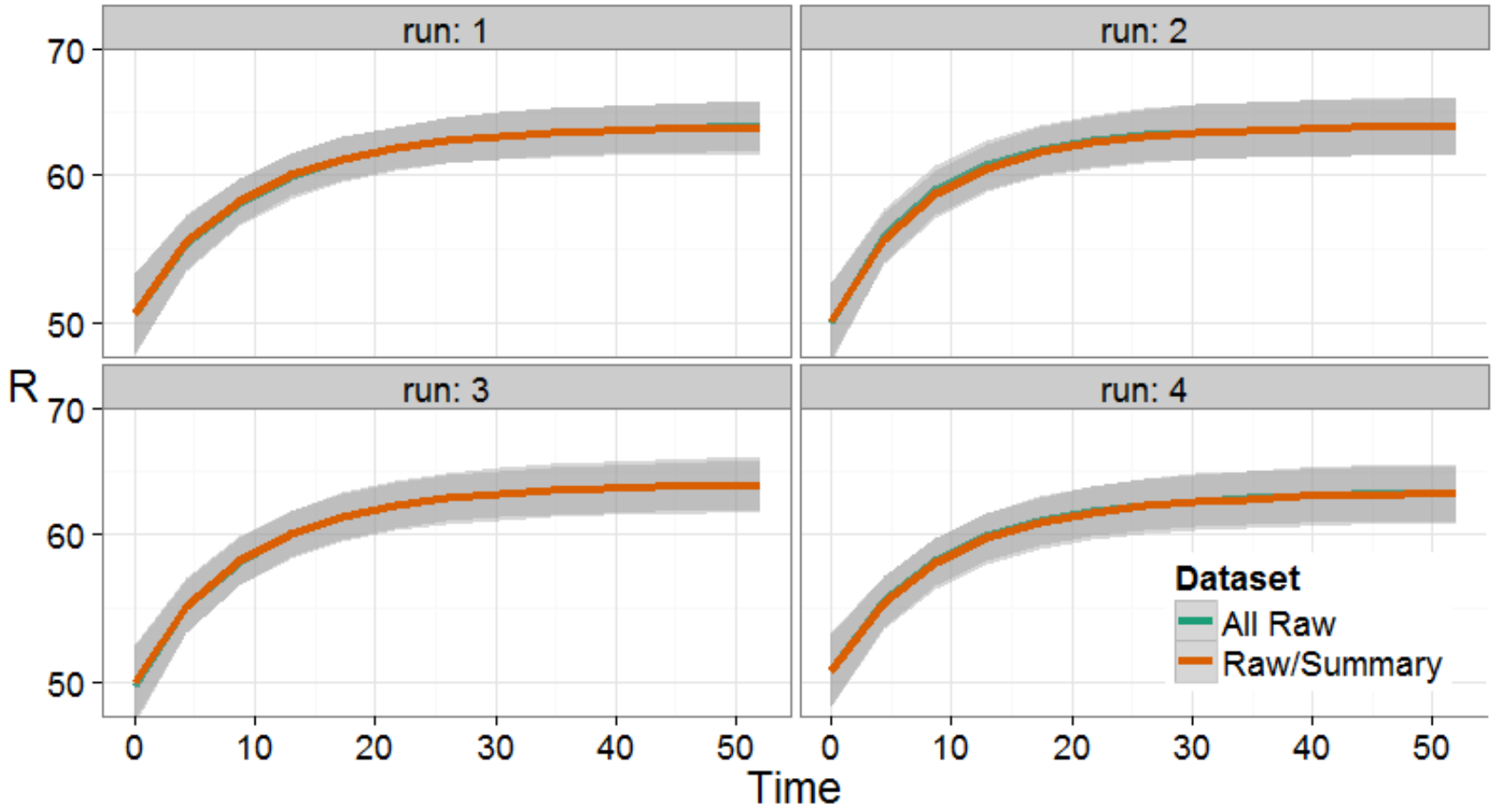
Shown is The All Raw Data Case



NOVARTIS

Comparison for the Prediction of the 95% CI Mean

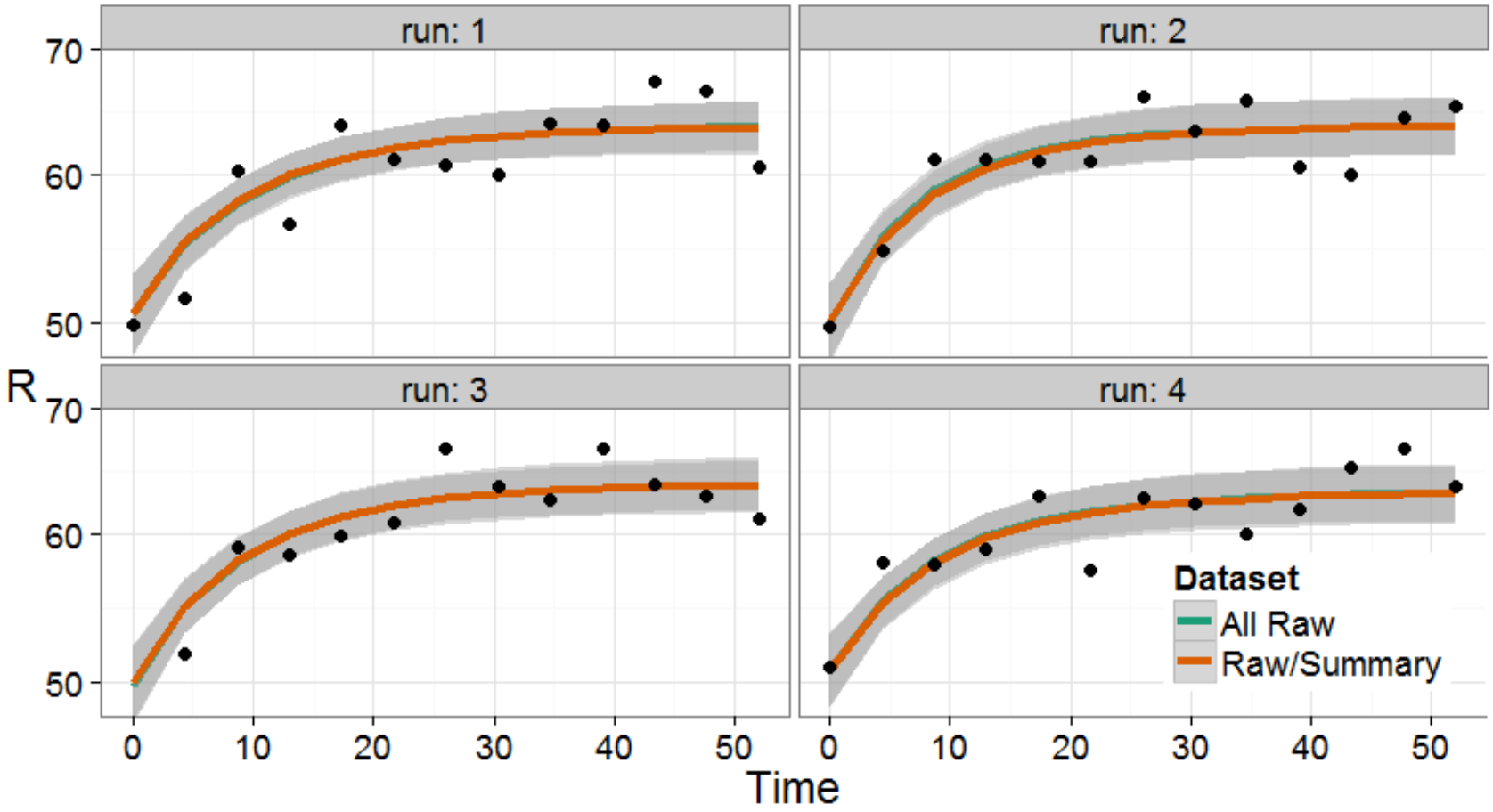
Shown is The All Raw Data Case with Raw/Summary Overlaid



NOVARTIS

Comparison for the Prediction of the 95% CI Mean

Shown is The All Raw Data Case with Raw/Summary Overlaid



NOVARTIS

Outlook: Hierarchical Expectation Propagation

HEP Approximation Promises Improved Robustness

- Inference problem assumed to be separate for ϕ and δ, i.e. *external data is not informative about* ϕ
- If external data is informative about ϕ then importance weights may become unstable
- **Hierarchical Expectation Propagation (HEP)** alleviates this by **data partitioning** in that it splits the approximation of the likelihood into parts, i.e. In raw / summary and updates in each iteration the ϕ and δ posterior sequentially.
  → EP-like algorithm

Conclusion

Utility of Non-Linear Hierarchical Models Greatly Expanded

- **Non-linear hierarchical** models offer great flexibility during study design, yet they are of limited use in situations with heterogeneous sources of information
  → Usually combination of raw and summary intractable

- Key assumptions to combine raw and summary data
  1. **Partially shared model**
  2. **Approximate patient likelihood with Normal probability model for external summary (central limit theorem justification)**
  3. **Efficiency ensured via suitable (iterative) proposal density; High efficiency of 20% to 40%**

- Generally applicable, fake-data cross-check advisable

Acknowledgements

Colleagues at Novartis

- David James
- Ramesh Sarangapani
- Beat Neuenschwander
- Simon Wandel
- Satrajit Roychoudhury

Co-Authors

- Andrew Gelman
- Bob Carpenter
- Daniel Lee
- Frederic Y. Bois
- Amy Racine

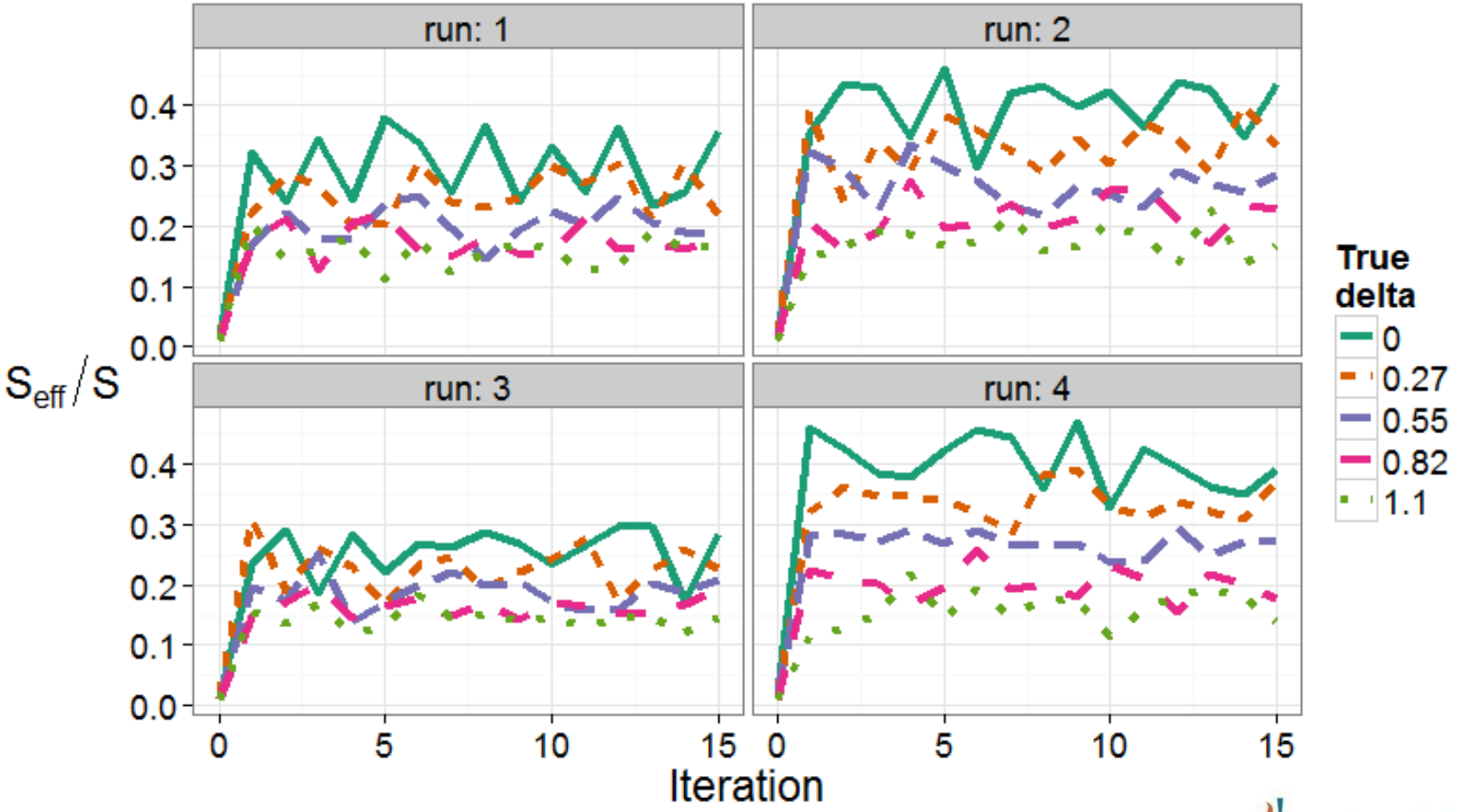
References

- Stan
  - Stan Dev Team 2015, Version 2.6. http://mc-stan.org.
  - Hoffman and Gelman, JMLR., 2014, Vol. 15, 1351
- Expectation Propagation
  - Minka, T. Proc. 17. Conf. On Uncert. In AI, 362-369, 2001
  - Gelman, A. et al., Technical report, Columbia University, 2014
- Importance Resampling
  - Rubin, Ann. Stat., 1984, Vol. 12, 1151
  - **Smith and Gelfand, Am. Stat., 1992, Vol. 46, No. 2**
  - Diggle and Gratton, J Roy. Stat. Soc., 1984, Series B 46:193

Efficiency vs Iteration for the Raw/Summary Case

Efficiency Variation due to Realization and Decrease with Larger δ

- Effective sample size
  $S_{eff}^{-1} = \sum_{s=1}^{S} w_s^{\dagger 2}$
- Few iterations for optimal performance needed



NOVARTIS