Bayesian Methods in Regulatory Science: Identifying Patient Subgroups with Positive Treatment Effects

Bradley P. Carlin

Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN

with many thanks to

Mr. Patrick Schnell, University of Minnesota
Drs. Qi Tang and Walter Offen, AbbVie, Inc.

ACDRS Workshop:
"Borrowing Strength from Accumulating Data"
Washington, DC / April 21, 2016

Overview of Current Research

- Adaptive Incorporation of Historical Data: Companies often eager to borrow strength from historical data, but proper amount is often subjective and controversial –
  - Any prior distribution favorable to the company's position risks Type I error inflation
  - Quality of historical data may vary widely, say by age or study type (RCTs, case-control, single-arm, observational, etc.)
- Possible solutions:
  - "Back out" the information content of the prior based on a predesignated upper bound on Type I error
  - Power Priors (Ibrahim and Chen, 2000): handy if degree of borrowing, α, can be predetermined; however, often computationally awkward to place a hyperprior on α
  - Commensurate Priors (Hobbs et al., 2011, 2012): degree of borrowing is determined in part by the "commensurability" (similarity) of the information of the historical and current data Example: For current and historical parameters θ and θ_0, $p(\theta|\theta_0, \eta) = N(\theta_0, \eta^{-1})$ with a gamma or "spike and slab" hyperprior on η (spike at a large η_0, slab over $0 < \eta < \epsilon$)

Important Applications

- Rare and pediatric diseases: Use of commensurate prior enables cautious use of historical data on a rare disease, or adult data for a pediatric drug or device approval (Gamalo et al., 2016, DIA Bayesian Working Group paper)
  - Children represent a large underserved population:
    - 80% of children are treated off-label $\Rightarrow$ safety, efficacy, and PK/PD of such drug therapies is unknown
    - drug development in children is often delayed or abandoned due to difficulty in running clinical trials
  - Ongoing work at Minnesota testing "Lorenzo's Oil" in adrenoleukodystrophy (ALD): PK/PD studies are underway; these results inform a subsequent Bayesian adaptive Phase IIa efficacy trial (Basu et al., 2015)
  - Also working on power and commensurate prior models for incorporating adult longitudinal data ($N_{adult} = 1137$) on both efficacy and safety of the drug cinacalcet in pediatric kidney disease ($N_{peds} = 40$)

Important Applications (cont'd)

- Combining Randomized and Nonrandomized Data: "Correct" the NR data using propensity score or other causal methods, then incorporate into the analysis using commensurate priors
- Also working to incorporate differential propensity weighting of patient-level data: For studies $s = 1, \cdots, S$, where $s = 1$ denotes the primary study, incorporate study $s > 2$ according to the "propensity" for its being included in study 1
  - Account for bias arising from inter-study heterogeneity as patient-level "weights" via a power prior $\Rightarrow$ permits integration of R and NR cohorts under the usual assumption that confounding is accounted for by the measurable covariates

Currently doing this in the context of an HIV/AIDS study ("FIRST") that featured a optional randomized substudy, so we have both randomized and nonrandomized groups that meet the same entry criteria (Zhao et al., 2015)

Control of Type I error in Regulatory Science

- Regulators tend to care much more about false positives (Type I error) than they do about false negatives:
  - Safety concerns: rofecoxib (Vioxx): a nonsteroidal anti-inflammatory drug (NSAID) prescribed to over 80M people for arthritis or other chronic pain; withdrawn from the market in 2014 over concerns about increased risk of heart attack and stroke associated with long-term, high-dosage use (annual sales at that time: $2.5B)
  - Efficacy concerns: flibanserin (Addyi; "female Viagra"): initial trial indicated an increase of one satisfying sexual encounter per month (baseline: 2 to 3/month); subsequent *JAMA Internal Medicine* meta-analysis of eight studies of 5900 women decreased the benefit to just one-half of an additional sexually satisfying encounter per month (annual sales: $11M)

Regulators often seek to control Type I error, and "pay" for this with (sometimes large) increases in Type II error (false negatives).

Companies may feel differently, especially in early discovery phase!

Identifying Interesting Patient Subgroups

- Clinical trials are traditionally designed to estimate the "overall" effect γ of a treatment T, e.g.,

$$\log \text{odds} = \alpha + \beta X + \gamma T$$

  where X is a *prognostic* covariate (say, age; gives information about the outcome *regardless* of treatment)

- BUT: Modern treatments don't work the same way for everyone (effect heterogeneity). Enhance model to include a *predictive* covariate Z

$$\log \text{odds} = \alpha + \beta X + (\gamma_0 + \gamma_1 Z) T$$

  So if $Z = 1$ for males and 0 for females, then the treatment effect is $\gamma_0 + \gamma_1$ for males, but γ_0 for females

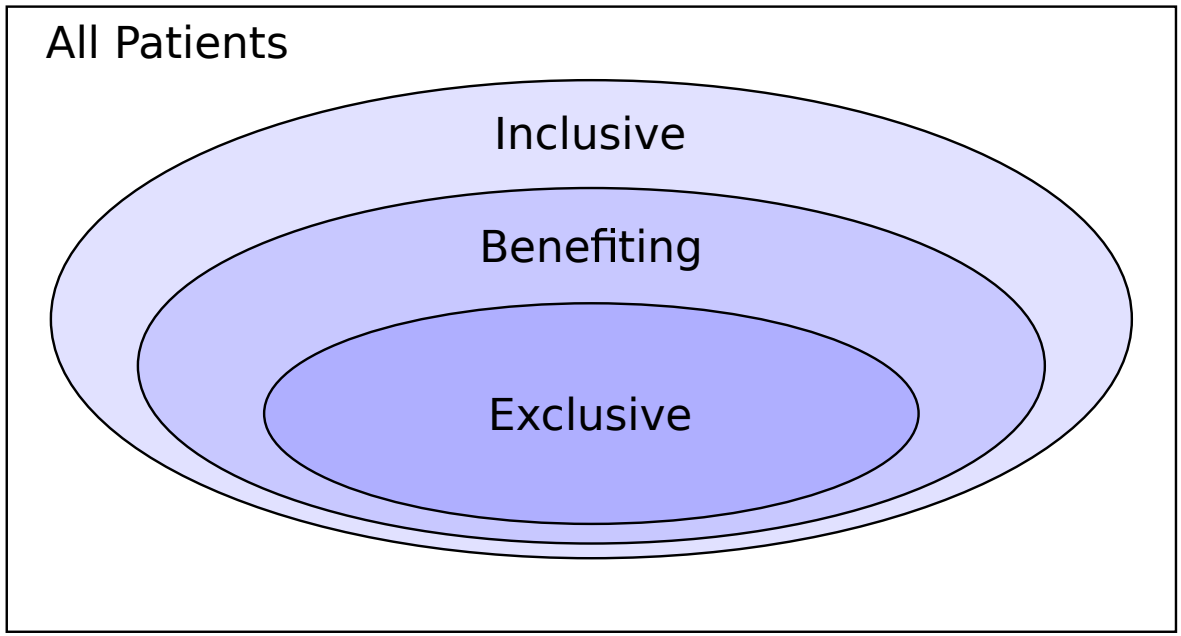
- Pharma companies no longer want to ask, "Does it work?"; they want to ask, "For whom does it work?"

Credible Subgroups

- **Exclusive credible subgroup** should contain only patients who benefit
- **Inclusive credible subgroup** should contain every patient who benefits

$$\text{Exclusive} \subseteq \text{Benefiting} \subseteq \text{Inclusive}$$

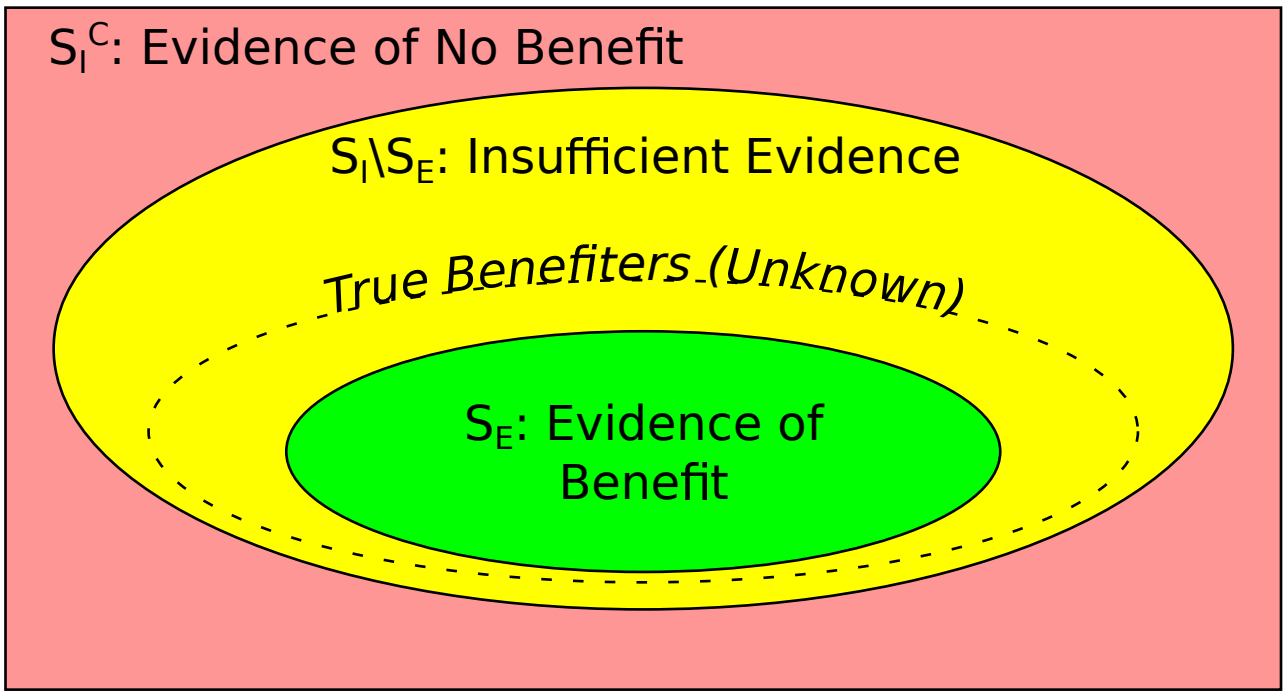
(Analogy with credible/confidence intervals: $L \leq \theta \leq U$)



Formal Definition of Credible Subgroups

If S_B is the benefiting subgroup, then the $(1-\alpha)$-level inclusive credible subgroup, S_I, and exclusive credible subgroup, S_E, are subsets of the population such that the posterior probability given the data D that $S_E \subseteq S_B \subseteq S_I$ is at least $1-\alpha$, i.e.

$$P(S_E \subseteq S_B \subseteq S_I | D) \geq 1-\alpha$$



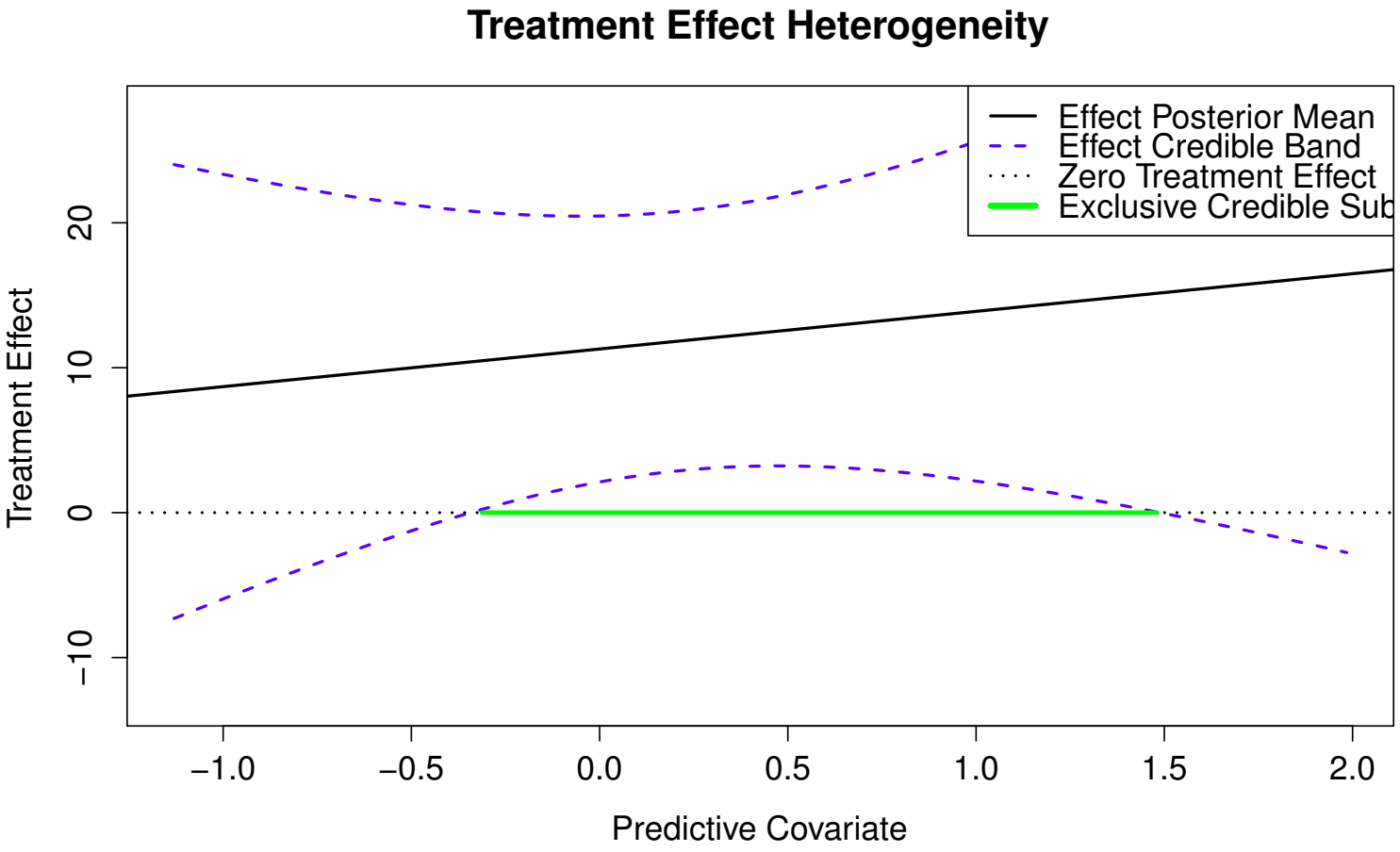
Credible Subgroups for Linear Models

- Suppose $E(Y_i|\mathbf{x}_i, \mathbf{z}_i, t_i) = \mathbf{x}_i'\boldsymbol{\beta} + t_i\mathbf{z}_i'\boldsymbol{\gamma}$
- We seek the subgroup of patients for which

$$\Delta(\mathbf{z}) = E(Y|\mathbf{x}, \mathbf{z}, t = 1) - E(Y|\mathbf{x}, \mathbf{z}, t = 0) = \mathbf{z}'\boldsymbol{\gamma} > \delta \ ,$$

  which we define as the benefiting subgroup.
- Possible implementation procedure:
  1. Find the $1 - \alpha$ highest posterior density (HPD) region for $\boldsymbol{\gamma}$
  2. If $\mathbf{z}_i'\boldsymbol{\gamma} > \delta$ for **all** $\boldsymbol{\gamma}$ in the HPD, then $\mathbf{z}_i$ is in S_E
  3. If $\mathbf{z}_i'\boldsymbol{\gamma} > \delta$ for **any** $\boldsymbol{\gamma}$ in the HPD, then $\mathbf{z}_i$ is in S_I
- Approximate frequentist guarantee under noninformative priors
- Works for entire (infinite) predictive covariate space; conservative on a restricted covariate space
- More generally, $\Delta(\mathbf{z})$ could be difference in log odds, etc.

Illustration in 1-d

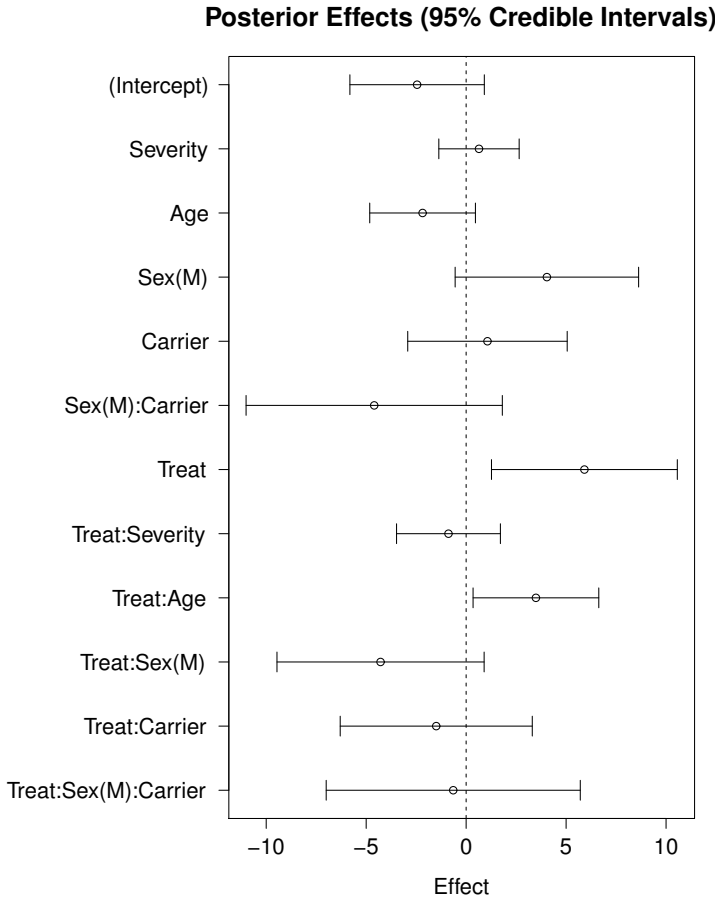


Regression to estimate personalized treatment effects, then *simultaneous* thresholding (here, for $\delta = 0$). Green region is S_E.

Example: Treatment for Alzheimer's Disease

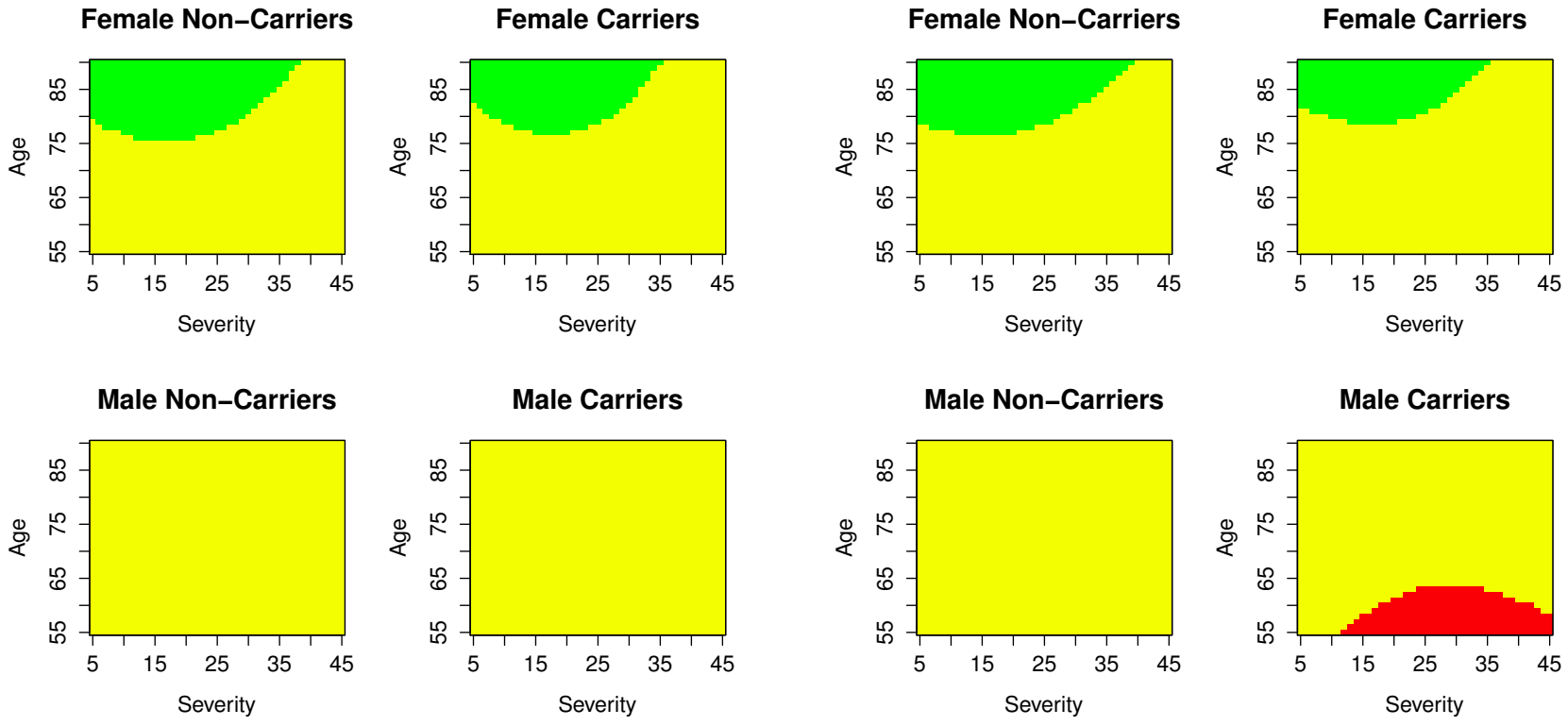
- Several covariates may be predictive
  - **Age:** 55–90 years
  - **Sex:** male and female
  - **Carrier:** carrier of ApoE4 allele
  - **Severity:** 5–45 in ADAS-Cog 11 score (lower is better)
- Response: **Change** in Severity (baseline $\rightarrow$ 24 weeks)
- Every covariate is treated as both prognostic and predictive (X_{ij} and Z_{ij} for patient i, covariate j)
- Assume response Y_i is normally distributed
- Vague priors for intercept, prognostic, and overall treatment effect
- Skeptical $N(0, 1)$ priors for treatment-covariate interactions
- Use 10,000 samples from the (multivariate t) posterior...

Regression Fit



Posterior summaries of regression parameters; continuous covariates are standardized. Only `Treatment` and `Treatment` $\times$ `Age` are "significant" (no multiplicity adjustment).

RCS Credible Subgroups



Left: 80% RCS credible subgroups with effect threshold $\delta = 0$
Right: 50% RCS credible subgroups with effect threshold $\delta = 2$

Color key: green, evidence of benefit; yellow, insufficient evidence; red, evidence of no benefit.

Current and Future Work

- Basic idea in Schnell et al. (2016a; this talk) enables multiplicity-protected subset selection in normal hierarchical linear models with a single endpoint
- Multiplicity-correcting methods (such as ours) maintain extremely high specificity at the expense of sensitivity; uncorrected methods do the opposite
- Extension to multiple endpoints (say, 2 efficacy and 1 safety) – requires a utility function, or some notion of admissibility
  - Broadly, a treatment is admissible $\Leftrightarrow$ there is no other treatment that is better w.r.t. every endpoint
  - Schnell et al. (2016b) rigorize this to weak and strong admissibility
- Extension to multiple treatments: complicates assessment of which comparisons we care about (i.e., count for multiplicity adjustment)

Papers Related to this Work

- Basu, C., Ahmed, M., Kartha, R.V., Brundage, R.C., Raymond, G.V., Cloyd, J.C., and Carlin, B.P. (2015). A hierarchical Bayesian approach for combining pharmacokinetic/pharmacodynamic modeling and phase IIa trial design in orphan drugs: treating adrenoleukodystrophy with Lorenzo's Oil. Submitted to *J. Biopharmaceutical Statistics*.
- Hobbs, B.P., Carlin, B.P., Mandrekar, S., and Sargent, D.J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, **67**, 1047–1056.
- Schnell, P.M., Tang, Q., Müller, P., and Carlin, B.P. (2016b). Credible subgroup inference for multiple treatments and multiple endpoints. Research report, Division of Biostatistics, University of Minnesota.
- Schnell, P.M., Tang, Q., Offen, W.W., and Carlin, B.P. (2016a). A Bayesian credible subgroups approach to identifying patient subgroups with positive treatment effects. To appear *Biometrics*.
- Zhao, H., Hobbs, B.P., Ma, H., Jiang, Q., and Carlin, B.P. (2015). Combining non-randomized and randomized data in clinical trials using commensurate priors. Submitted to *Health Services and Outcomes Research Methodology* (ICHPS 2015 special issue).