Using Bayesian methods to analyze data from N-of-1 trials

Christopher H Schmid

Brown University School of Public Health
Department of Biostatistics

Recent Advances in Meta-Analysis Symposium
Göttingen, Germany
29 August 2024

Outline

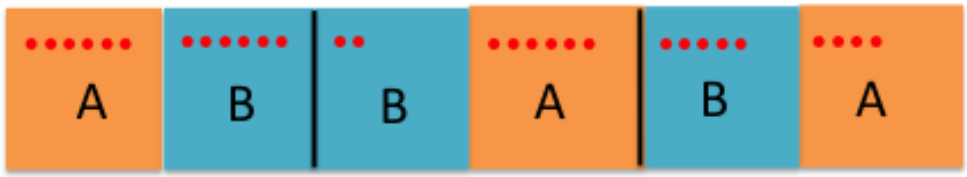
- N-of-1 Design
- Examples
- Model for One Individual
- Combining N-of-1 trials
- Study Networks

Single Case Experimental Designs

- Personalized protocol for each participant
- Contains multiple treatment periods in which different intervention given
- Repeated measurements taken in each period
- Many types:
  - Multiple Baseline (stepped wedge)
  - Alternating Treatment
  - Changing Criterion
  - Withdrawal/Reversal (multiple crossover)
  - N-of-1 (randomized multiple crossover)
- Can measure individual treatment efficacy
- Can combine single-case trials in multilevel structure
  - Estimate average effects
  - Assess heterogeneity
  - Improve individual estimates

Nikles et al. (2015) The Essential Guide to N-of-1 Trials in Health

N-of-1 Design



- Single patient multiple period blocked crossover trials to estimate individual treatment effects
- Multiple measurements per period
- Potential missing data
- Compare measurements in A periods with those in B periods
- May need washout to control for carryover

Kravitz and Duan (2014), Design and Implementation of N-of-1 Trials: A User's Guide, AHRQ Publication

Indications

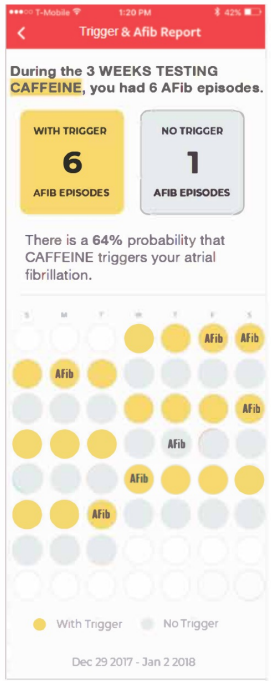
- Substantial therapeutic uncertainty about treatment
- Heterogeneous treatment effects
- Stable chronic condition
- Short-acting treatments with rapid ramp-up
- Negligible persistence of treatment effect (no carryover)
- Outcome expected to return to baseline after each period
- Measurable, easily collected outcomes

Motivation for Personalized (N-of-1) Trials

- Traditional clinical trials focused on average participant often sampled from specialized population
- Implement research in community and local practice
- Expand experimental research into underrepresented populations and everyday life
- Facilitate individual decision making
- Estimate individual treatment effects
- Assess heterogeneity of treatment effects

N-of-1 Trials: Personalized protocols

- Design own study
- Select own (multiple) outcomes
- Select own treatments
- Determine data collection process
- Use mobile platforms: design setup, randomization, data collection, reminders, analysis



Key Design Elements For Type 1 N-of-1 Trials

- Assigning Treatments
  - Randomization within crossovers
  - Systematic counterbalanced design (AB/BA)
  - Alternating Treatments
- Allocation concealment
- Blinding
- Replication to assess within and between period variability
  - Number of study periods, number of measurements per period
  - Patients may not finish their protocol
- Carryover of treatment effect usually biases toward null
  - Designed washout period lengthens study and may be impractical/counterproductive or unethical
  - Analytic washout reduces sample size
- Multiple outcomes
  - May want to weight to make decision

Examples of N of 1 Studies

Condition	Pts	Xovrs	Tx length	Outcome	Comparison
Fibromyalgia	58	3	6 wk	FIQ	AM/ AM + FL
Chronic Pain	98	2-4	1-2 wk	Various	Various
Inflammatory Bowel	54	2	8 wk	Various	Strict/relaxed diet
Atrial Fibrillation	446	3	1 wk	Episodes	Trigger/none
Daily mood	447	3	5 day	Various	Various

- Many other application in literature now

PREEMPT Study: Design

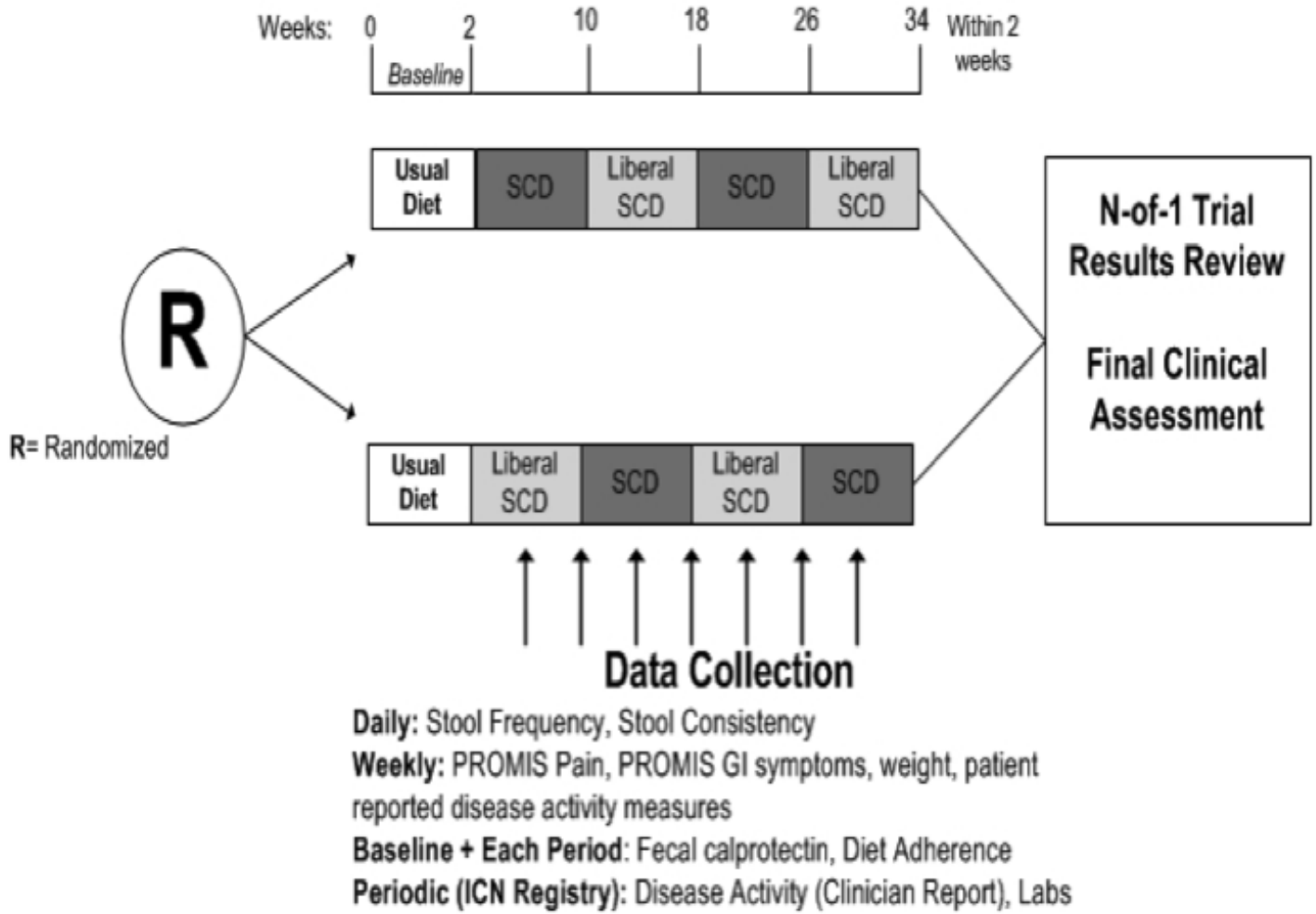
- Compares N-of-1 trials versus usual care for treating adults with chronic musculoskeletal pain (215 randomized)
- Compare outcomes 3, 6, 12 months after baseline for N-of-1 vs. usual care
- Outcomes: Pain, Quality of life, Participatory decision making, Satisfaction, Trust, Adherence

Kravitz et al., *JAMA Internal Medicine* 178:1368-1378, 2018

PREEMPT N-of-1 Study Arm Protocol

- Develop mobile application to conduct N-of-1 trials (108 patients)
- Compare 2 interventions within each patient
  - 1-2 week treatment periods
  - Cycle of 2 periods (2 to 4 weeks long, AB or BA)
  - Study of 2-4 cycles (4-16 weeks)
- Self-reported daily outcomes: pain, fatigue, drowsiness, sleep problems, cognitive function, constipation
- Choice of treatments by patient/clinician (pharmaceuticals, opiates, non-pharmaceutical)
- Patients continue on concomitant treatments

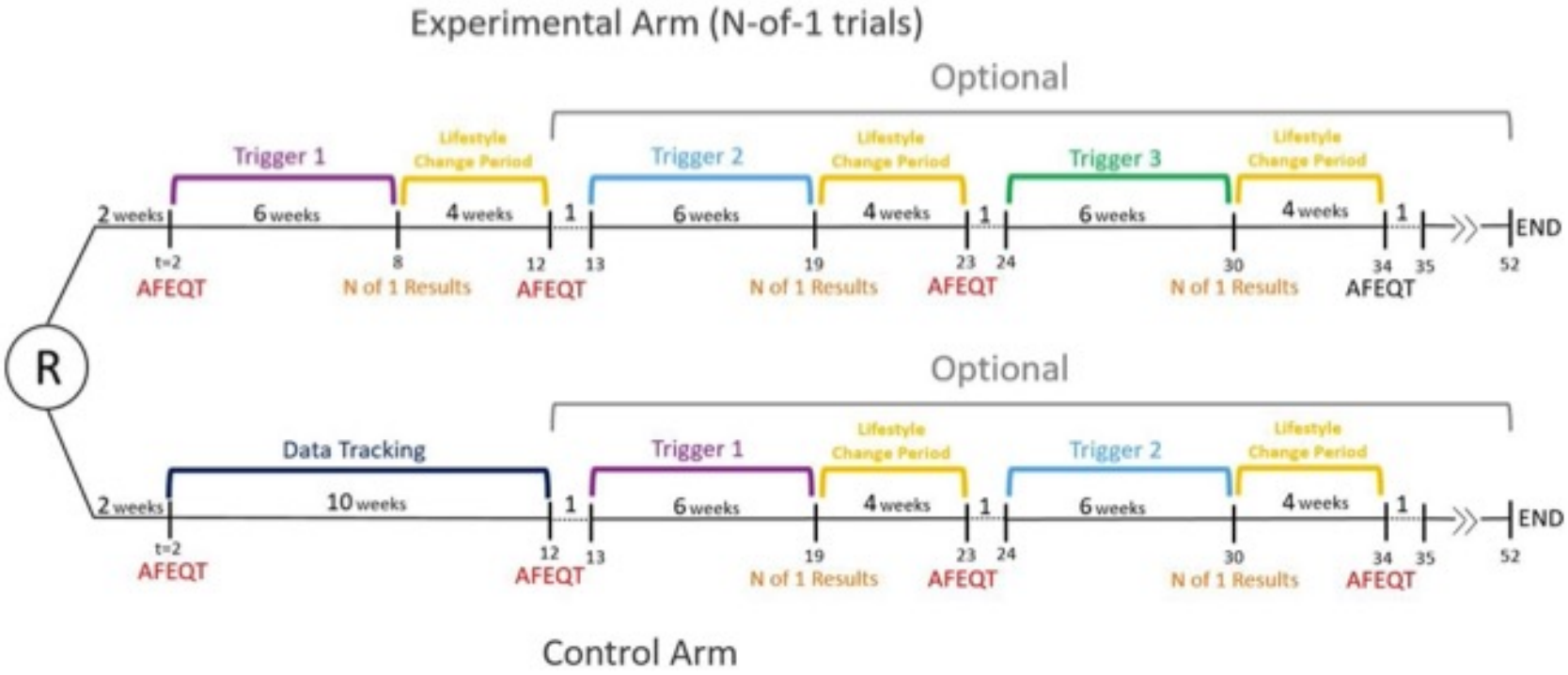
PRODUCE Study



- 54 N-of-1 trials comparing Specific Carbohydrate Diet (SCD) to liberalized SCD and baseline for 7-18 year old patients with Crohn's Disease (CD), Ulcerative Colitis (UC), or Indeterminate Colitis (IC)

I-STOP-AF Study Design

- Find triggers (e.g., alcohol, caffeine, exercise) for discrete episodes of periodic atrial fibrillation (AF)
- Patient networks including HealtheHeart Study and StopAfib.org



Marcus et al., *JAMA Cardiology* 7:167–174, 2022

Statistical Analysis

- Structured time series with treatment factor
- Descriptive analysis with graphics (visual inspection)
- Basic statistics comparing one treatment to another (e.g., paired t test)
- Modeling to address
  - Time-trends
  - Time-varying treatment effects
  - Autocorrelation
  - Carryover
  - Treatment Interactions

Model for Single N-of-1 Trial: Treatment Only

$$y_j = \mu + \delta z_j + \epsilon_j; \qquad j = 1, 2, \ldots, J$$

$$\epsilon_j \sim N(0, \sigma^2)$$

y_j: measurement j for outcome y

z_j: treatment indicator

Single Trial Model: Treatment + Linear Time Trend

$$y_j = \mu + \delta z_j + \beta t_j + \epsilon_j; \qquad j = 1, 2, \ldots, J$$

$$\epsilon_j \sim N(0, \sigma^2)$$

y_j: measurement j for outcome y
z_j: treatment indicator
t_j: time of j^{th} measurement

Single Trial Model: Tx + Linear Trend + Correlated Error

$$y_j = \mu + \delta z_j + \beta t_j + \epsilon_j; j = 1, 2, \ldots, J$$

$$\epsilon_j = \rho_e \epsilon_{j-1} + u_j$$

$$u_j \sim N(0, \sigma^2)$$

y_j: measurement j for outcome y
z_j: treatment indicator
t_j: time of j^{th} measurement

Under stationarity

$$\epsilon_j \sim N(0, \sigma^2/(1-\rho_e^2))$$

Marginally,

$$Y_j \sim N\left(\mu + \delta z_j + \beta t_j, \sigma^2/(1-\rho_e^2)\right)$$

Extensions to Basic N-of-1 Models

- Add interaction of treatment with time

$$y_j = \mu + \delta z_j + \beta t_j + \gamma z_j \times t_j + \epsilon_j$$

- Use nonlinear time (e.g., cubic spline)

$$y_j = \mu + \delta z_j + F(t_j) + \epsilon_j$$

$$F(t_j) = \boldsymbol{B}(t_j) \text{ where } \boldsymbol{B}(t_j) = \sum_{m=1}^{M} \gamma_m B_m(t_j) \text{ is a spline}$$

- Use period blocks as factor

$$y_j = \mu + \delta z_j + \sum_i \gamma_i p_{ij} + \epsilon_j$$

  where p_{ij} is an indicator for block i at time j
- Carryover
- Lagged outcome

Rationale for Using Bayesian Models

- Personalized nature of decision
- Desire to incorporate external information (patient, clinician)
- Posterior distribution of difference between treatments
- Joint posterior distribution for composite statements
- Lack of sufficient data for frequentist to return 'significant' result
- Can handle missing data as parameters (imputation)
- Incorporate informative prior information
- Combining multiple N-of-1 studies together gives average treatment effect and better individual treatment effects through borrowing of strength

Schmid and Yang. Bayesian models for n-of-1 trials. Harvard Data Science Review, 2022

Automating Model Fitting

- Real-time environment based on mobile applications requires automated analyses that can be returned quickly without need for statistician to check results
- Implemented using smartphones in real-time environment
- Use Bayesian model implemented via MCMC in R and JAGS so that probabilities can be returned
- Can use informative or non-informative priors
- May want to compare different prespecified models using Bayesian model fitting criteria (size of effect, DIC, posterior predictive checks)
- May wish to introduce some lag time for statistician to check results

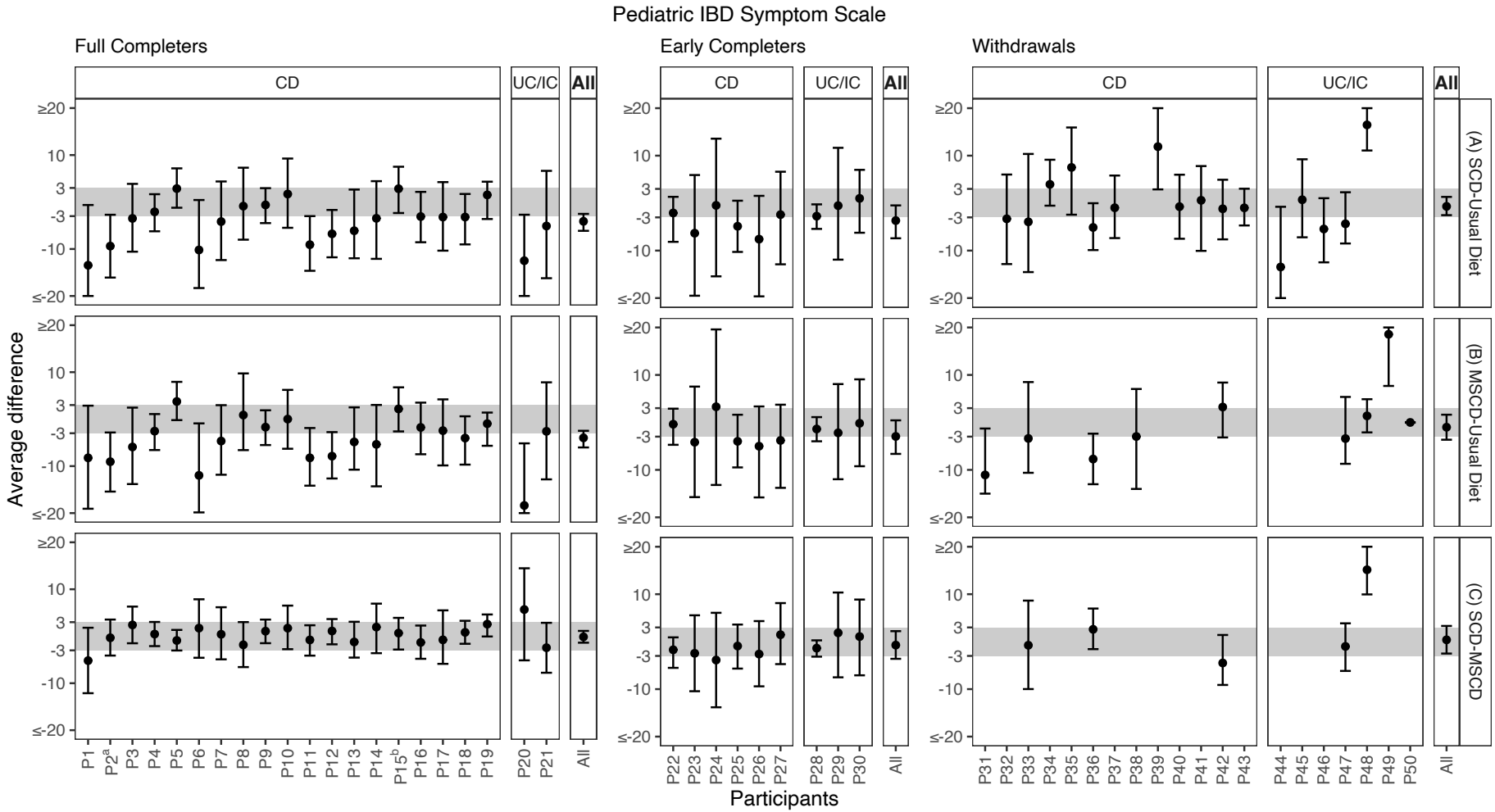
Choosing Prior Distribution

- Priors for mean parameters like μ, β, γ and δ usually noninformative, e.g., $N(0, 10^6)$
- Noninformative prior like $U(-1, 1)$ on correlation may be too weak if likelihood ohas little information about correlation parameters
- Posterior inferences most sensitive to choice of prior for variance σ^2
- Must use distribution with support only on positive numbers
  - Bounded uniforms
  - Folded (half) normal or t
  - Lognormal

Analysis of PRODUCE Diet Trials

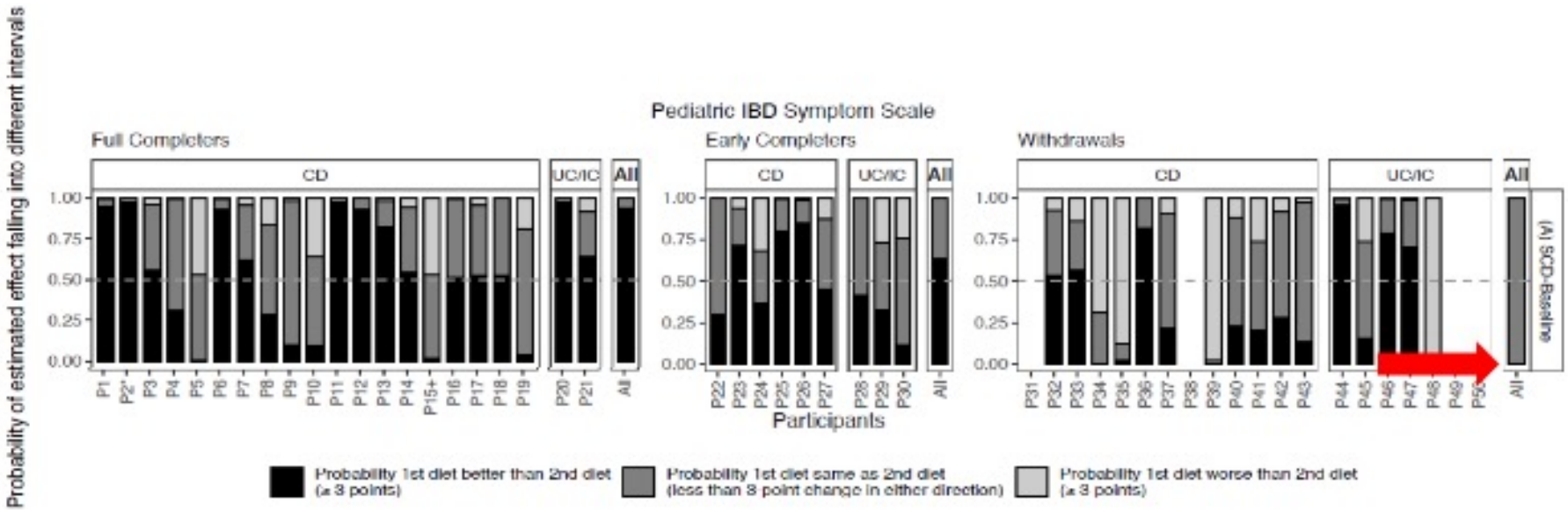
- Use weekly Pediatric IBD Symptom Scale, reported as a T-score (standardized mean of 50 and standard deviation of 10)
- Clinically important change is defined as a 3-point change in the scale in either direction
- Did not analyze first weekly measurement in any experimental diet period to account for carryover
- Among 54 randomized participants, 21 completed the full four-period sequence, 9 completed the study early after a single crossover (two periods), and 24 withdrew during the first or second period before completing both diets
- 12/21 full completers, 5/9 early completers, and 4/24 withdrawals classified as responders on SCD compared to UD
- 12/21 full completers, 4/9 early completers, and 1/24 withdrawals classified as responders on MSCD compared to UD

PRODUCE GI Symptoms: Individual Results



PRODUCE GI Symptoms: Individual Results

IBD Symptom Scale: SCD vs. Usual Diet (Withdrawals)



*P9 is child response data
+P12 started on incorrect diet (not randomized diet)

PRODUCE Study Conclusions

- 39% of patients who completed a full trial had high probability of improvement
- Heterogeneity in response
- Low probability that SCD was better than MSCD (individually and on average)
- Most electing to continue dietary intervention chose to stay on less restrictive MSCD
- 61% of patients withdrew or completed the study early
- N-of-1 trials are useful for determining not just whether dietary therapy with the SCD or MSCD works on average, but also for whom it works

Combining N-of-1 Studies

- Can treat each N-of-1 trial as a study and combine via meta-analysis (multilevel models)
- Get average estimate of treatment efficacy
- Get improved estimates for individuals by borrowing strength
- Compromise between population estimate (complete pooling) and individual's observed results (no pooling)
  - Weighted to observed if low variation or many crossovers
  - Weighted to pooled (or subgroup) if little information for individual
- Can include covariates for heterogeneity and subgroups
- Can include terms for carryover, correlation
- If treatments differ, can form network

Zucker et al. Combining Single Patient (N-of-1) Trials to Estimate Population Treatment Effects and to Evaluate Individual Patient Responses to Treatment. Journal of Clinical Epidemiology 50:401-410, 1997

Combining N-of-1 Studies via Meta-Analysis

Observation at time j from person (study) i:

$$y_{ij} = \mu_i + \delta_i z_{ij} + (\text{Trend} + \text{Carryover}) + \epsilon_{ij}$$

$$\epsilon_{ij} = \rho_{e_i} \epsilon_{i(j-1)} + u_{ij}$$

$$u_{ij} \sim N(0, \sigma_i^2)$$

- $\delta_i \sim N(d, \sigma_\delta^2)$ are individual treatment effects
- d is average treatment effect
- Fixed or random effect for μ_i
- Random effect ρ_{e_i} for correlation
- Separate/common within-study variances $\sigma_i^2 = \sigma^2$
- Can also add covariates into model either for adjustment or as treatment interactions
- Can have these modify the individual parameters such as δ_i, μ_i

Combining N-of-1 Studies via Meta-Analysis

- If intercepts are treated as random, then model can incorporate correlation $\rho_{\mu\delta}$ between intercepts and treatment effects
- Within-individual correlations ρ_{e_i} more complicated to model because they are typically skewed and bounded
- Can use inverse hyperbolic tangent (Fisher z) transformation $z_{e_i} = \frac{1}{2} \ln\left(\frac{1+\rho_{e_i}}{1-\rho_{e_i}}\right)$ and assume that

$$z_{e_i} \sim N(z_e, \sigma^2_{z_e})$$

- Then $\rho_{e_i} = \exp(2z_{e_i} - 1)/(1 + \exp(2z_{e_i}))$
- Can estimate average effects of hyperparameters and updated estimates of individuals parameters

Two Bivariate Model Forms

Intercept-Slope form

$$Y_{ij} = \mu_i + \delta_i Z_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} = \rho_e \epsilon_{ij-1} + u_{ij}$$

$$u_{ij} \sim N(0, \sigma^2)$$

$$\begin{pmatrix} \mu_i \\ \delta_i \end{pmatrix} \sim N\left[\begin{pmatrix} m \\ d \end{pmatrix}, \begin{pmatrix} \sigma_\mu^2 & \rho_{md}\sigma_\mu\sigma_\delta \\ \rho_{md}\sigma_\mu\sigma_\delta & \sigma_\delta^2 \end{pmatrix} \right]$$

Two Bivariate Model Forms

Intercept-Slope form

$$Y_{ij} = \mu_i + \delta_i Z_{ij} + \epsilon_{ij}$$
$$\epsilon_{ij} = \rho_e \epsilon_{ij-1} + u_{ij}$$
$$u_{ij} \sim N(0, \sigma^2)$$

$$\begin{pmatrix} \mu_i \\ \delta_i \end{pmatrix} \sim N\left[\begin{pmatrix} m \\ d \end{pmatrix}, \begin{pmatrix} \sigma_\mu^2 & \rho_{md}\sigma_\mu\sigma_\delta \\ \rho_{md}\sigma_\mu\sigma_\delta & \sigma_\delta^2 \end{pmatrix} \right]$$

Separate means form

$$y_{ij} = (1 - z_{ij})\mu_{i0} + z_{ij}\mu_{i1} + \epsilon_{ij}$$
$$\epsilon_{ij} = \rho_e \epsilon_{ij-1} + u_{ij}$$
$$u_{ij} \sim N(0, \sigma^2)$$

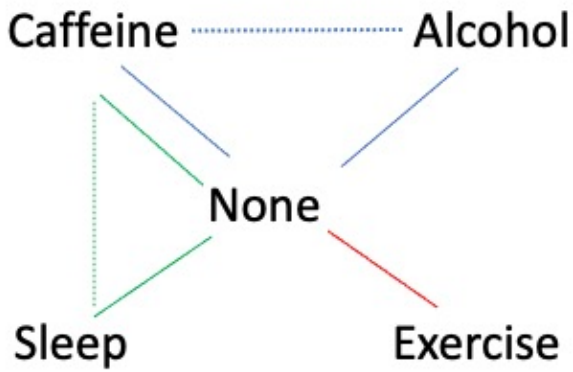
$$\begin{pmatrix} \mu_{i0} \\ \mu_{i1} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho_{01}\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix} \right]$$

Two Bivariate Model Forms

- 1:1 mapping between parameters of two models
- When $\sigma_0 = \sigma_1$,
  - either $\sigma_\delta = 0$,i.e., no heterogeneity
  - or $\rho_{md} = -\sigma_\delta / 2\sigma_\mu$, i.e., slope and intercept are negatively correlated
- ρ_{md} only non-negative if $\rho_{01} \geq 1/k$, i.e., only if correlation between group means is large or variance of treated much greater than variance of controls
- Variety of values for ρ_{01} compatible with negative ρ_{md}
- $U(-1, 1)$ prior may work better for ρ_{01} than for ρ_{md}
- Instead might want to use prior for ρ_{md} weighted toward negative values

I-STOP-AF Study: Network Meta-Analysis

- Set of trigger comparisons forms network of triggers
- Compare triggers indirectly through common control using network meta-analysis
- Combine all trials for a participant into one trial with several arms, one for each trigger and one combining no trigger periods
- OR Treat each participant's trials as separate and introduce a common participant effect into a mixed model

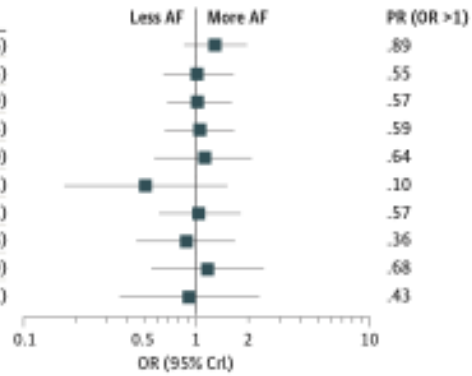


ISTOP-AFIB Network Meta-Analysis

Figure 2. Network Meta-analyses Odds Ratios for Self-reported AF During Trigger Exposure vs Avoidance in Intention to Treat and Per Protocol Combining All n-of-1 Trials Throughout the Study

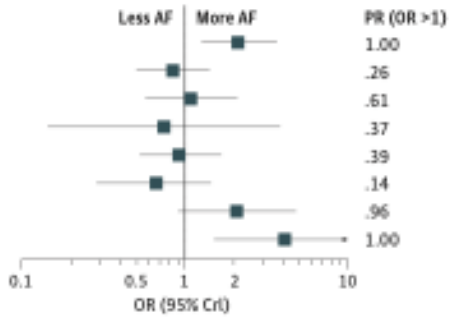
A Intention to treat

Trigger type	On No.	On Total No.	Off No.	Off Total No.	OR (95% CrI)	PR (OR >1)
Alcohol	164	923	154	924	1.30 (0.85-1.96)	.89
Caffeine	301	1045	310	1047	1.03 (0.65-1.65)	.55
Lack of sleep	125	750	110	726	1.03 (0.68-1.59)	.57
Exercise	162	794	135	761	1.05 (0.67-1.63)	.59
Dehydration	51	375	43	410	1.12 (0.58-2.10)	.64
Cold food and drink	22	106	29	111	0.49 (0.17-1.51)	.10
Lying on the left side	101	424	115	445	1.05 (0.62-1.81)	.57
Large meals	106	348	106	348	0.88 (0.46-1.68)	.36
Diet	46	227	43	226	1.21 (0.54-2.49)	.68
Custom	45	228	35	209	0.91 (0.36-2.31)	.43



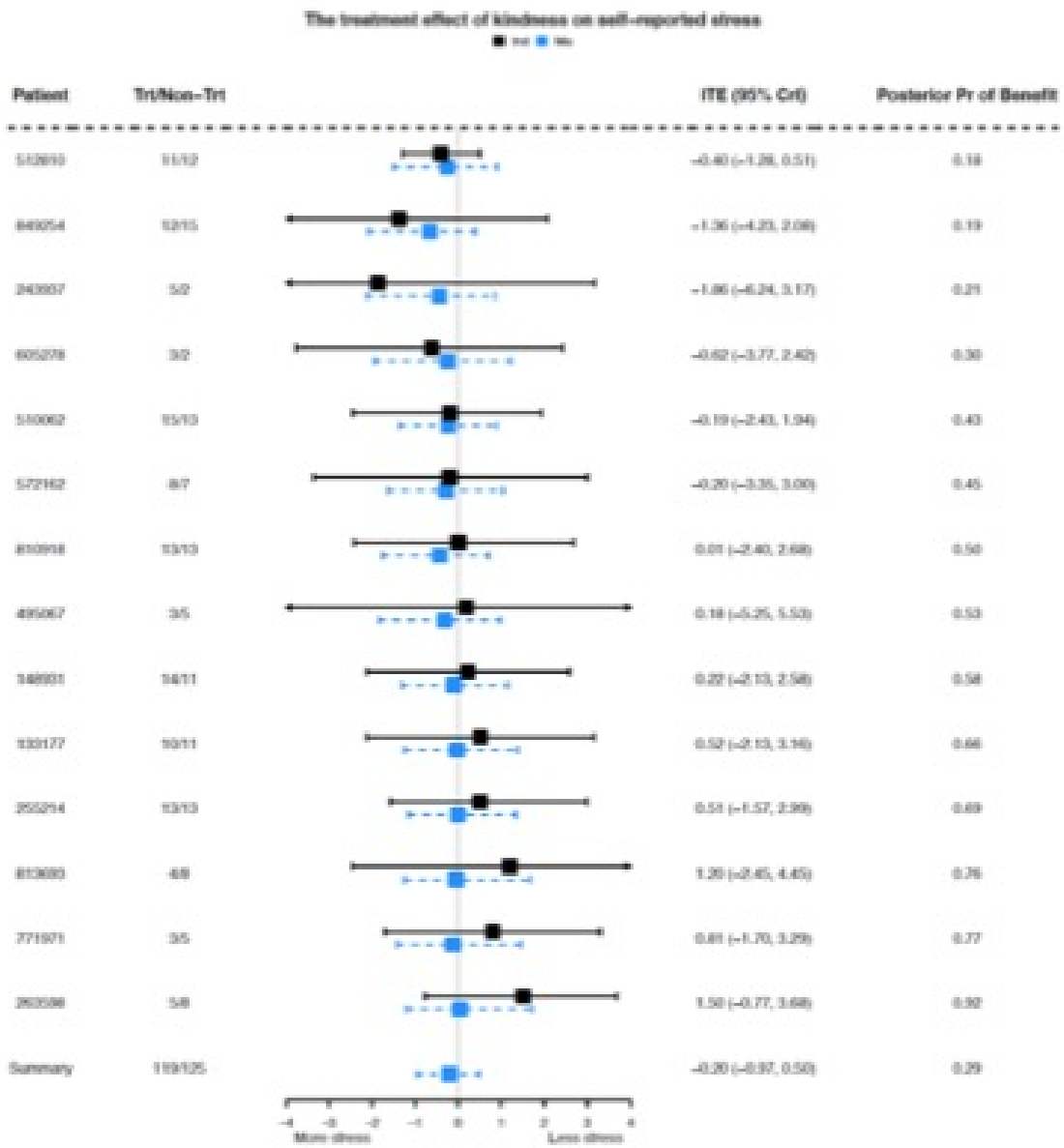
B Per protocol

Trigger type	On No.	On Total No.	Off No.	Off Total No.	OR (95% CrI)	PR (OR >1)
Alcohol	148	578	141	913	2.15 (1.27-3.61)	1.00
Caffeine	291	841	268	885	0.84 (0.51-1.40)	.26
Exercise	88	508	99	448	1.10 (0.58-2.06)	.61
Cold food and drink	6	52	6	56	0.74 (0.14-3.80)	.37
Lying on the left side	106	462	108	403	0.92 (0.52-1.66)	.39
Large meals	67	243	143	448	0.67 (0.30-1.45)	.14
Diet	49	180	34	224	2.11 (0.91-4.73)	.96
Custom	47	197	30	213	4.09 (1.49-11.58)	1.00



- Similar to results from separate meta-analysis of each trigger

Shrinkage Estimation

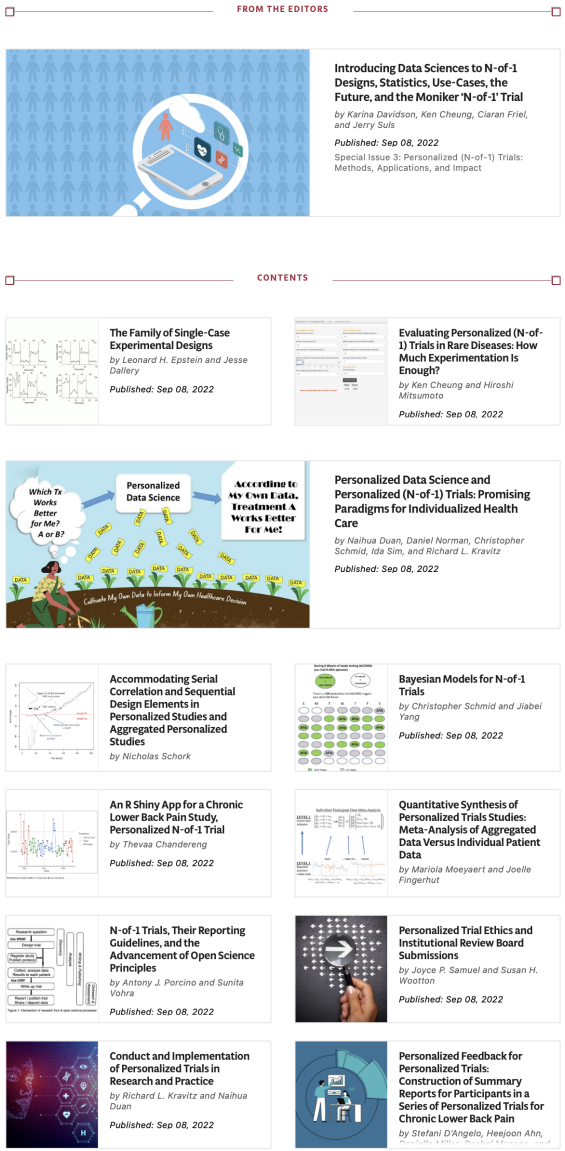


Conclusions

- N-of-1 trials can be a useful tool to personalize treatments and discover individual effects
- They require some infrastructure, but tools are under development
- Need to determine how best to return information to individual participants
- Trials from different individuals can be combined to inform population effects and to improve predictions for individuals
- Can be part of a larger network to inform treatment comparisons

References: Harvard Data Science Review Special Issue

https://hdsr.mitpress.mit.edu/specialissue3



FROM THE EDITORS

Introducing Data Sciences to N-of-1 Designs, Statistics, Use-Cases, the Future, and the Moniker 'N-of-1' Trial
by Karina Davidson, Ken Cheung, Ciaran Friel, and Jerry Suls
Published: Sep 08, 2022
Special Issue 3: Personalized (N-of-1) Trials: Methods, Applications, and Impact

CONTENTS

The Family of Single-Case Experimental Designs
by Leonard H. Epstein and Jesse Dallery
Published: Sep 08, 2022

Evaluating Personalized (N-of-1) Trials in Rare Diseases: How Much Experimentation Is Enough?
by Ken Cheung and Hiroshi Mitsumoto
Published: Sep 08, 2022

Personalized Data Science and Personalized (N-of-1) Trials: Promising Paradigms for Individualized Health Care
by Naihua Duan, Daniel Norman, Christopher Schmid, Ida Sim, and Richard L. Kravitz
Published: Sep 08, 2022

Accommodating Serial Correlation and Sequential Design Elements in Personalized Studies and Aggregated Personalized Studies
by Nicholas Schork

Bayesian Models for N-of-1 Trials
by Christopher Schmid and Jiabei Yang
Published: Sep 08, 2022

An R Shiny App for a Chronic Lower Back Pain Study. Personalized N-of-1 Trial
by Thevaa Chandereng
Published: Sep 08, 2022

Quantitative Synthesis of Personalized Trials Studies: Meta-Analysis of Aggregated Data Versus Individual Patient Data
by Mariola Moeyaert and Joelle Fingerhut

N-of-1 Trials, Their Reporting Guidelines, and the Advancement of Open Science Principles
by Antony J. Porcino and Sunita Vohra
Published: Sep 08, 2022

Personalized Trial Ethics and Institutional Review Board Submissions
by Joyce P. Samuel and Susan H. Wootton
Published: Sep 08, 2022

Conduct and Implementation of Personalized Trials in Research and Practice
by Richard L. Kravitz and Naihua Duan
Published: Sep 08, 2022

Personalized Feedback for Personalized Trials: Construction of Summary Reports for Participants in a Series of Personalized Trials for Chronic Lower Back Pain
by Stefani D'Angelo, Heejoon Ahn,

Thank you!