Adaptive Designs in Surveys and Clinical Trials: Similarities, Differences, and Opportunities for Cross-fertilization[1]

Thomas A. Louis, PhD

Department of Biostatistics
Johns Hopkins Bloomberg SPH
tlouis@jhu.edu

Expert Statistical Consultant
Center for Drug Evaluation & Research
U.S. Food & Drug Administration
Thomas.Louis@fda.hhs.gov

[1] Presented at the 6^{th} workshop: Advances in Adaptive and Responsive Survey Design: From Theory to Practice, 4-5 November 2019, U. S. Census Bureau.

Goals & Outline[2]

Goals

- Highlight opportunities for technology transfer
- Identify a few research ideas

Outline

- Overview of survey and clinical trial adaptations
- Examples of Survey and of Clinical Trial adaptations
- Survey $\longleftrightarrow$ Clinical
- Coda: Care is needed

[2] Presentation based in part on: Rosenblum M, Miller P, Reist B, Stuart EA, Thieme M, Louis TA (2019). Adaptive Design in Surveys and Clinical Trials: Similarities, Differences, and Opportunities for Cross-Fertilization. *J. Roy. Statist. Soc., Ser. A*, 182: 963–982. DOI: 10.1111/rssa.12438.

Goals & Outline[2]

Goals

- Highlight opportunities for technology transfer
- Identify a few research ideas

Outline

- Overview of survey and clinical trial adaptations
- Examples of Survey and of Clinical Trial adaptations
- Survey $\longleftrightarrow$ Clinical
- Coda: Care is needed

Some displayed details are FYI and won't be discussed

[2] Presentation based in part on: Rosenblum M, Miller P, Reist B, Stuart EA, Thieme M, Louis TA (2019). Adaptive Design in Surveys and Clinical Trials: Similarities, Differences, and Opportunities for Cross-Fertilization. *J. Roy. Statist. Soc., Ser. A*, 182: 963–982. DOI: 10.1111/rssa.12438.

Types of Adaptation (a subset)

In Trials

Stop early: for efficacy, futility or harm (group sequential designs)

Modify criteria: enrollment, dose, sample size, follow-up time, randomization probabilities or endpoints

Target recruitment: to 'enrich' with potential responders to treatment

Adjust randomization: to over-populate the apparently better treatment

Re-randomize: participants with poor outcomes to another treatment; 'Sequential, Multiple Assignment Randomized Trials' (SMART)

Types of Adaptation (a subset)

In Trials

Stop early: for efficacy, futility or harm (group sequential designs)

Modify criteria: enrollment, dose, sample size, follow-up time, randomization probabilities or endpoints

Target recruitment: to 'enrich' with potential responders to treatment

Adjust randomization: to over-populate the apparently better treatment

Re-randomize: participants with poor outcomes to another treatment; 'Sequential, Multiple Assignment Randomized Trials' (SMART)

In Surveys

Stop early: for 'efficacy' (sufficient data) or futility (little potential for more)

Dynamically: target, enrich and suppress

Efficiently allocate: data collection resources

Mode-switch: start with the web; delay ?? days before sending hard copy

Modify timing: or frequency of contact attempts

Change incentives: for participating or responding

Augment R-factors: to include effects of ultimate analysis

Bureaucratic Traction in Clinical Trials

Official Guidance

- The European Medicines Agency in 2007 and the U. S. FDA in 2016 and 2018
  - For all FDA guidances and more, visit, https://www.fda.gov/drugs/guidance-compliance-regulatory-information

Question

- Is there, or should there be, similar guidance from AAPOR or other organization; possibly, from the ASD group?

Bureaucratic Traction in Clinical Trials

Official Guidance

- The European Medicines Agency in 2007 and the U. S. FDA in 2016 and 2018
  - For all FDA guidances and more, visit, https://www.fda.gov/drugs/guidance-compliance-regulatory-information

Question

- Is there, or should there be, similar guidance from AAPOR or other organization; possibly, from the ASD group?

Innovation at the FDA

As displayed in the *Federal Register* notice on August 29, 2018, FDA is conducting a Complex Innovative Trial Design (CID) Pilot Meeting Program to support the goal of facilitating and advancir the use of complex adaptive, Bayesian, and other novel clinical trial designs. The CID Pilot Meeting Program fulfills a performance goal agreed to under PDUFA VI, included as part of the FDA Reauthorization Act of 2017.

Survey $\longrightarrow$ clinical trial

Monitor representativeness and improve it by targeted enrollment or follow-up

- To improve internal validity: compare baseline variables of respondents to those of overall sample, and target intensive follow-up (double-sampling) of non-responders to increase balance/representativeness
- To improve external validity: monitor how representative the enrolled participants are of the target population and selectively increase efforts to enrol underrepresented groups
- Use R-indicators to measure balance/representativeness, and determine which baseline variables contribute most to it

Collect and use paradata to improve retention and protocol compliance

- Number of attempts needed to schedule visit
- Arrival time (late or early)
- Number of questions answered and time on each question in interviews
- Clinician observations on participant (dis)satisfaction with study experience
- Use paradata to predict participant retention and protocol compliance
  - Then, identify whom to target with interventions that encourage participation and/or protocol compliance

Clinical Trial $\longrightarrow$ Survey

A chartered Data Monitoring Committee

- Constitute a chartered, arms-length committee with the appropriate expertise and freedom from conflict of interest that meets at regular intervals, including pre-study initiation

Clinical

- Called a Data Monitoring Committee (DMC), a Data and Safety Monitoring Committee (DSMB), ...
  - Monitors study conduct (enrollment, data timeliness and quality), participant safety, treatment efficacy or futility
  - Makes recommendations to the study sponsor

Survey

- The DMC/DSMB could evaluate the frame and monitor:
  - Survey conduct (enrollment, data timeliness and quality)
  - Implementation of adaptive decisions (timing, frequency, contact mode for non-respondents)
  - Respondent burden (e.g., from multiple contacts)
  - Disclosure avoidance measures

Clinical Trial $\longrightarrow$ Survey

A chartered Data Monitoring Committee

- Constitute a chartered, arms-length committee with the appropriate expertise and freedom from conflict of interest that meets at regular intervals, including pre-study initiation

Clinical

- Called a Data Monitoring Committee (DMC), a Data and Safety Monitoring Committee (DSMB), ...
  - Monitors study conduct (enrollment, data timeliness and quality), participant safety, treatment efficacy or futility
  - Makes recommendations to the study sponsor

Survey

- The DMC/DSMB could evaluate the frame and monitor:
  - Survey conduct (enrollment, data timeliness and quality)
  - Implementation of adaptive decisions (timing, frequency, contact mode for non-respondents)
  - Respondent burden (e.g., from multiple contacts)
  - Disclosure avoidance measures

Is a survey DMC/DSMB worth considering?

Clinical Trial $\longrightarrow$ Survey

Sequential Multiple Assignment Randomized Trial (SMART) designs

- In each wave, participants are **randomized** to different contact modes, intensities or incentives to respond

Goals (somewhat in competition)

- **Conduct** a good survey
- **Learn** which sequences are most effective in producing sample balance, decreasing cost or decreasing survey duration[3]

In surveys

- Identify optimal (at least very good) sequential treatment rule within strata of auxiliary variables using methods of Murphy (2003); Robins (2004); van der Laan & Luedtke (2015)
- For example, target non-respondents most likely to increase sample representativeness (e.g., R–indicator) at lowest cost

Issue

- Requires modeling, and so vulnerable to model misspecification
  - Necessary for (almost) all adaptive designs

[3] Dworak and Chang (2015) randomized non-respondents in the Health and Retirement Survey to different sequences of $$ and persuasive messages.

SMART Surveys

Get Smart

- Specify mode sequences, then randomize to sequences or sequentially randomize to learn what works well
- If embedded in a real survey, make sure to maintain survey quality
  - **Balance learning and doing**

Notation (FYI)

$m_k =$ Planned mode sequence, e.g., $m_1 =$ internet, $m_2 =$ web, $m_3 =$ CATI, $\ldots$, m_K

- The m_k don't have to be unique, and 'mode' can have components
- 'internet:(no inducement)' and 'internet:inducement' are different modes

$Z \in \{1, 2, \ldots, K, K+1\}$ indicates the position in the sequence that generated the response ($Z = K+1$ indicates 'no response')

$m_Z =$ the mode that produced the response

- In reality full sequence up to and including m_Z is 'the mode'

$\tilde{Y} =$ The true, underlying value, assumed mode-independent

$Y =$ Reported value–depends on $\tilde{Y}$ and can depend on mode and mode sequence

$\mathbf{X} =$ Covariates

Monitoring Representativeness: necessary inputs

See[4,5] for Meng's cautions on lack of representation

Sampling frame (under-utilized in clinical and field studies)

- (Joint) distributions of a variety of attributes
- Benchmarking to frame and sample totals
- A high-quality sampling frame empowers effective adaptation

And, a subset of

- Mode-specific response time 'event curves'
- Propensity models for response, occupied unit, ...
  - Logistic or 'logic' regression, CART, random forests, ...
- Cost & Quality metrics
- Measures of statistical information

[4] Meng's discussion of Keiding&Louis (2016)

[5] Meng (2018). Statistical Paradises and Paradoxes in Big Data (I): Law of Large Populations, Big Data Paradox, and the 2016 Presidential Election. *Annals of Applied Statistics*, 12: 685–726.

Monitoring & Adjusting Representativeness

- Imbalance/balance indicators (Särndal, 2008, 2011; Särndal and Lundström, 2010; Lundquist and Särndal, 2013) and R-indicators (Schouten et al., 2009, 2011) identify,
  - Attributes that drive variation in response propensities and support adaptation by evaluating which subgroups are over/under represented
- Goals resonate with enriching a clinical trial

The sample R-indicator

- ρ_i is the estimated (possibly adjusted) response propensity for group i

$$R(\boldsymbol{\rho}) = 1 - 2\sqrt{\frac{1}{N-1}\sum_{1}^{N}(\rho_i - \bar{\rho})^2}$$

- $R(\boldsymbol{\rho}) = 1$ indicates that the sample is fully representative
  - Keiding & Louis[6,7] note that imbalance doesn't imply lack of representativeness

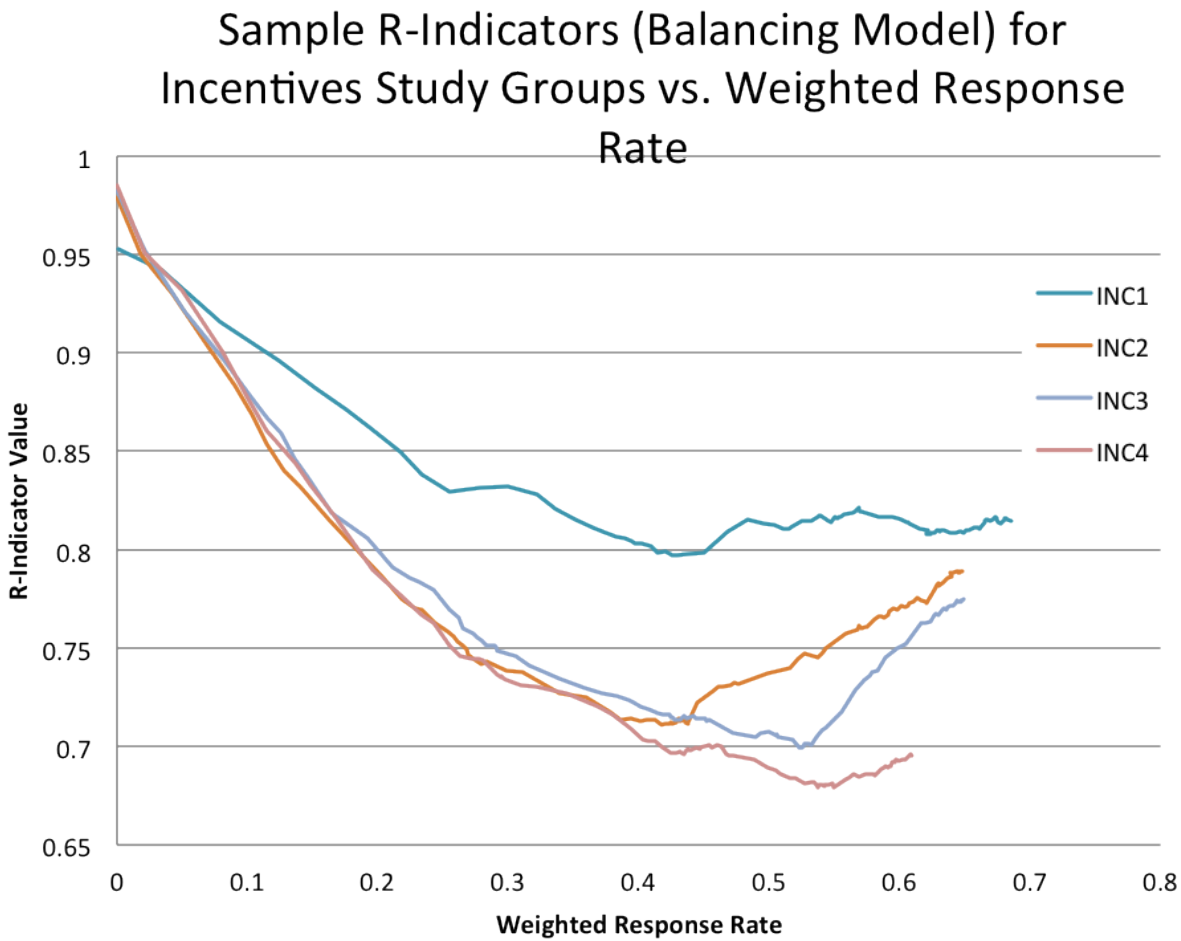
---

[6] Keiding N, Louis TA (2016). Perils and potentials of self-selected entry to epidemiological studies and surveys (with discussion and response). *J. Roy. Statist. Soc., Ser. A*, 179: 319–376.

[7] Keiding N, Louis TA (2018). Web-based Enrollment and other types of Self-selection in Surveys and Studies: Consequences for Generalizability. *Annual Review of Statistics and Its Application*, 5: 25–47.

Comparison of incentive approaches in the The National Survey of College Graduates[8]

- 4 separate surveys each using a different set of incentives, but with the same attributes used in the propensity model



[8] Thanks to Ben Reist

Partial, unconditional, R-indicators

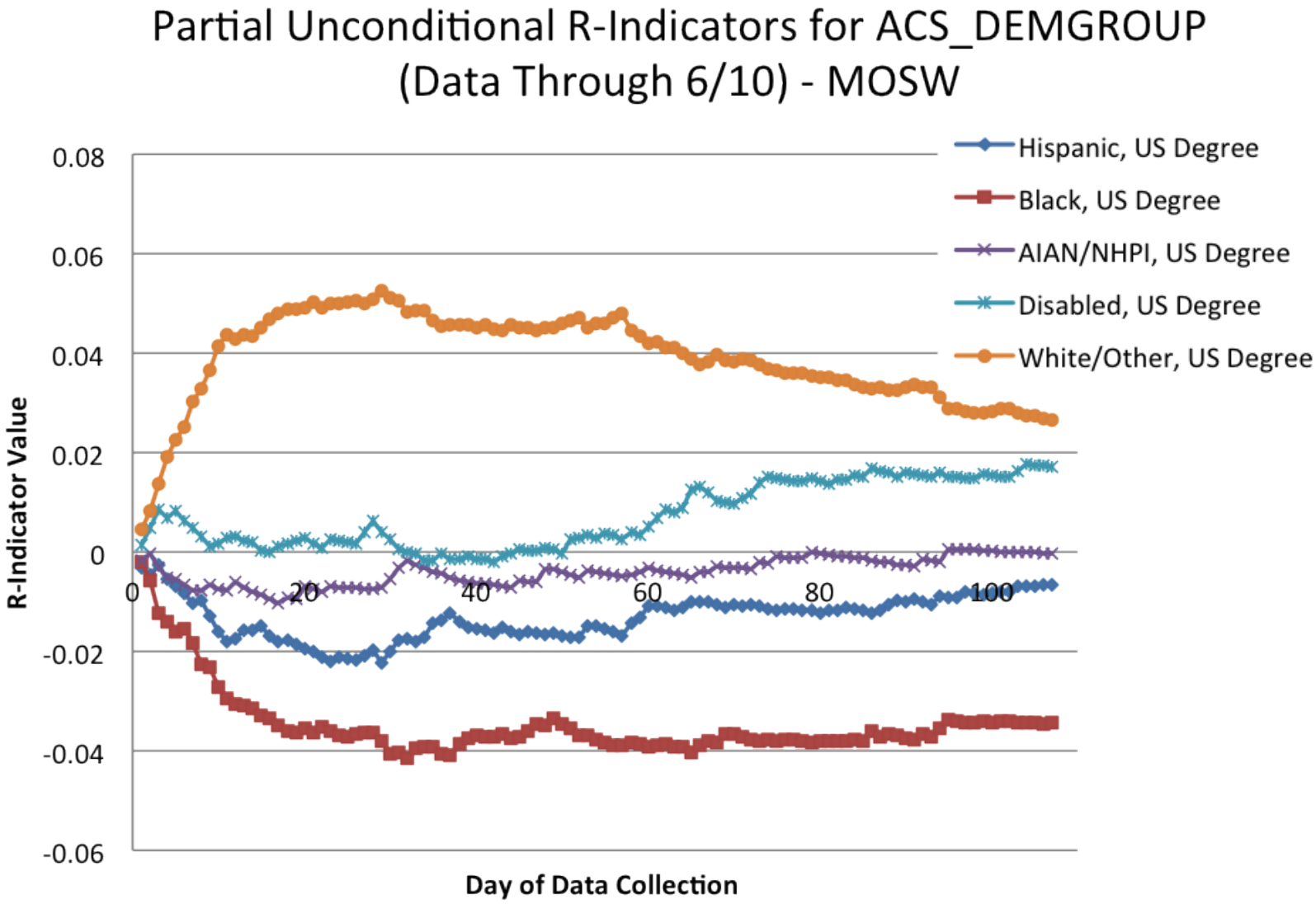
- Identify subgroups that are over/under represented
- Use the information to target cases; encourage or not encourage
- Adapt by switching modes, incentives, etc.
- With ρ_k the estimated (possibly adjusted) response propensity for group $X = k$, $\boldsymbol{\rho}$ the vector of indicators, and $\bar{\rho}$ the (weighted) mean, the unconditional R-indicator is

$$R_u(X = k, \boldsymbol{\rho}) = \left(\frac{N_k}{N_+}\right)^{\frac{1}{2}} (\rho_k - \bar{\rho})$$

- $R_u = 0 \Rightarrow$ balance

NSCG Data Monitoring Example

Could use a similar plot for clinic or subgroup representation



Moving beyond R-indicators

- Analysis of survey data can/should include, re-weighting, imputation, modeling, ...
- Survey cost is also a consideration
- So, include these in an adaptation criterion
- High-level view:
  - $\nu = T$: the last day of data collection
  - **Objective function:** Performance(T) = MSE or other quality metric
  - **Backward induction:** find the next-phase adaptation that maximizes,

$$E\left\{\text{Performance(T)} \mid \text{current data}^{(\nu)}, \text{adaptation}^{(\nu)}\right\}$$

    in one or many steps
- Bayesian structuring is almost essential

Moving beyond R-indicators

- Analysis of survey data can/should include, re-weighting, imputation, modeling, . . .
- Survey cost is also a consideration
- So, include these in an adaptation criterion
- High-level view:
  - $\nu = T$: the last day of data collection
  - **Objective function:** Performance(T) = MSE or other quality metric
  - **Backward induction:** find the next-phase adaptation that maximizes,

$$E\left\{\text{Performance(T)} \mid \text{current data}^{(\nu)}, \text{adaptation}^{(\nu)}\right\}$$

    in one or many steps
- Bayesian structuring is almost essential

Notation (FYI)

$\phi_k =$	frame fraction for sub-population k
$n_k^{(\nu)} =$	sample size for sub-population k at survey day ν
$f_k^{(\nu)} =$	$n_k^{(\nu)} / n_+^{(\nu)}$ = sample fraction for sub-population k at survey day ν
$(\mathbf{f}^{(\nu)}, \mathbf{n}_+^{(\nu)}, \boldsymbol{\phi}) =$	data at day ν
$T =$	last day of data collection
$(\mathbf{f}^{(\mathbf{T})}, \mathbf{n}_+^{(\mathbf{T})}, \boldsymbol{\phi}) =$	analysis data set
$\mathrm{M}(\mathbf{f}^{(\mathbf{T})}, \mathbf{n}_+^{(\mathbf{T})}, \phi \mid \text{analysis}) =$	Performance(T)

Clinical Trials: Allocation on Outcome

Bayesian Structuring $\approx$ Louis[9,10]

- Treatments T_1 and T_2, means $(\mu_1, \mu_2) \sim G$
- Sequential Probability Ratio Test (SPRT) stopping based on the likelihood-ratio (L_{mn}) after m responses on T_1 and n on T_2

  Continue if $0 < A < L_{mn} < B < \infty$
- Frequentist type I and II errors are controlled, even with adaptation
- With an equipoise (50/50) prior,
  $\pi_{mn} = pr(\mu_1 > \mu_2 \mid \text{data}) = L_{mn}/(1 + L_{mn})$

---

[9] Louis TA (1975). Optimal allocation in sequential tests comparing the means of two Gaussian populations. *Biometrika*, 62: 359–369.

[10] Louis TA (1977). Sequential allocation in clinical trials comparing two exponential survival curves. *Biometrics*, 33: 627–634.

Clinical Trials: Allocation on Outcome

Bayesian Structuring $\approx$ Louis[9,10]

- Treatments T_1 and T_2, means $(\mu_1, \mu_2) \sim G$
- Sequential Probability Ratio Test (SPRT) stopping based on the likelihood-ratio (L_{mn}) after m responses on T_1 and n on T_2

  Continue if $0 < A < L_{mn} < B < \infty$
- Frequentist type I and II errors are controlled, even with adaptation
- With an equipoise (50/50) prior,
  $\pi_{mn} = pr(\mu_1 > \mu_2 \mid \text{data}) = L_{mn}/(1 + L_{mn})$
- Select an imbalance bound: $0.5 \leq \phi < 1.0$
- If a large mean is good, allocate to keep,

$$\phi = 1: \qquad \frac{m}{m+n} \approx \pi_{mn}$$

$$\text{general } \phi: \qquad m/(m+n) \approx \phi\pi_{mn} + (1-\phi)(1-\pi_{mn})$$

- Optimizes a trade-off between total sample size and # on the inferior treatment
- The strategy is likely relevant to survey optimization

---

[9] Louis TA (1975). Optimal allocation in sequential tests comparing the means of two Gaussian populations. *Biometrika*, 62: 359–369.

[10] Louis TA (1977). Sequential allocation in clinical trials comparing two exponential survival curves. *Biometrics*, 33: 627–634.

Biometrika (1975), Simulation Results

Gaussian responses; T_1 is better

- M_ϕ and N_ϕ are expected sample sizes
- **Raw Cost:** excess total sample size $= (M_\phi + N_\phi) - (M_{0.5} + N_{0.5})$
- **Raw Benefit:** reduced assignment to the inferior treatment $= N_{0.5} - N_\phi$

$100\phi \rightarrow$	50	70
M_ϕ	78	127
N_ϕ	78	57
$M_\phi + N_\phi$	156	184
Raw Cost	0	**28**
Raw Benefit	0	**21**

Biometrika (1975), Simulation Results

Gaussian responses; T_1 is better

- M_ϕ and N_ϕ are expected sample sizes
- **Raw Cost:** excess total sample size $= (M_\phi + N_\phi) - (M_{0.5} + N_{0.5})$
- **Raw Benefit:** reduced assignment to the inferior treatment $= N_{0.5} - N_\phi$

$100\phi \rightarrow$	50	70
M_ϕ	78	127
N_ϕ	78	57
$M_\phi + N_\phi$	156	184
Raw Cost	0	**28**
Raw Benefit	0	**21**

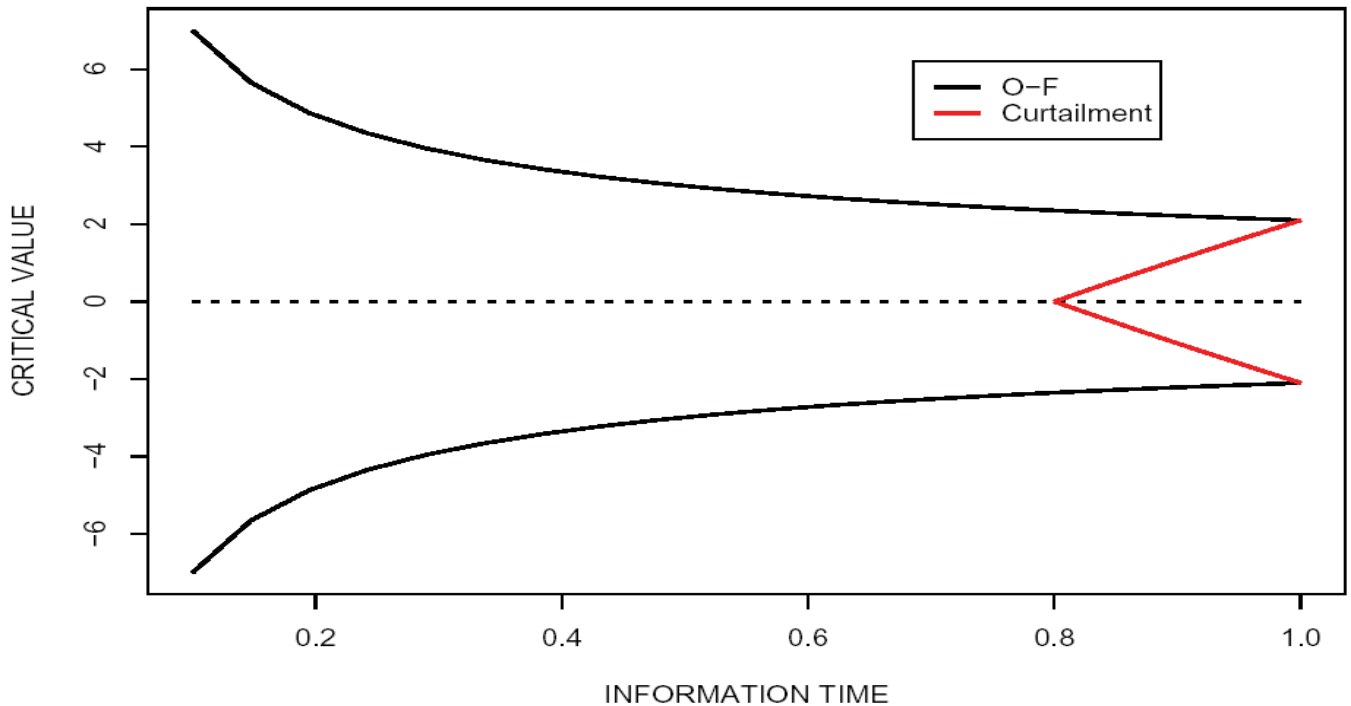
- Trade-offs: There is no free lunch
- Gain relative to 50/50: $\left(\frac{\phi}{1-\phi}\right) \times$ Benefit - Cost


I CAN'T BELIEVE SCHOOLS ARE STILL TEACHING KIDS ABOUT THE NULL HYPOTHESIS.
I REMEMBER READING A BIG STUDY THAT CONCLUSIVELY DISPROVED IT YEARS AGO.

From xkcd

Clinical trial stopping rules[11]

O'BRIEN-FLEMING
TWO-SIDED, CURTAILED



[11] DeMets DL, Friedman LM, Furberg CD (2006). *Data Monitoring in clinical trials.* New York: Springer.

Bayesian Monitoring: The BLOCK HF trial[12]

- Intention-to-treat analysis
- Adaptive Bayesian design with a maximum of 1200 patients
  - Two interim analyses with rules for sample size re-estimation, for stopping enrollment, and for terminating follow-up
- The safety stopping rule was based on the posterior probability of an increased risk of primary endpoints in patients with BiV pacing relative to RV pacing
- Terminating enrollment or follow-up was based on the predictive probability of

  $PP_0 = \text{pr(achieving the primary objective @ 12 mths fu| data, prior)}$
  $PPR = \text{pr(futility @ 12 mths fu | data, prior)}$

  projected to when all patients had been followed for at least 12 months
- Low information priors
- Substantial simulations to evaluate properties, including frequentist performance

[12] Curtis et al. (2013). Biventricular Pacing for Atrioventricular Block and Systolic Dysfunction. *NEJM*, 368: 1585–1593.

BLOCK-HF decision table

PP_0 = pr(achieving the primary objective @ 12 mths fu| data, prior)

PPR = pr(futility @ 12 mths fu | data, prior)

Decision Boundaries					
	Conclude objective is met and stop study early	Conclude that sample size is sufficient to continue	Determine that sample size is insufficient but elect not to increase sample size	Conclude that sample size must be increased in increments of 175	Stop study for safety
First Interim Analysis	$PP_0 > 0.99$	$0.90 \leq PP_0 \leq 0.99$	$PRR > 0.9$	$PP_0 < .90$ and $PRR \leq 0.9$	P ($\theta >$ 0|data,prior) $\geq$ 0.90
Sample Size Re-estimation Phase	N/A	$0.90 \leq PP_0$	$PRR > 0.9$	$PP_0 < .90$ and $PRR \leq 0.9$	N/A
Second Interim Analysis	$PP_0 > 0.99$	If neither the outcome in column 2 nor the outcome in column 6 occurs, then the study will continue with the current sample size.			P ($\theta >$ 0|data,prior) $\geq$ 0.90

Survey Stopping Rule[13]

- When is there sufficient information to stop conducting interviews?
- The 'stop and impute rule'

  $\hat{\theta}_{now}$: Use currently collected data, augmented by imputation of missing values
- The 'project rule'

  $\hat{\theta}_{future}$: Collect a specified number of additional interviews, and then augment by imputation of missing values
- Specify a discrepancy (ϵ) and an uncertainty (γ), then if a prediction model indicates that

$$\text{pr}\left(\mid \hat{\theta}_{\textbf{now}} - \hat{\theta}_{\textbf{future}} \mid > \epsilon \right) < \gamma,$$

  stop and use $\hat{\theta}_{now}$

Similar to futility assessment in a clinical trial

[13] Wagner, J. and Raghunathan, T. E. (2010) A new stopping rule for surveys. *Statist. Med.*, 29: 1014–1024.

Extension of Wagner & Raghunathan[14]

- Have information on n_1 of the n units in the sampling frame
- y_i is the observed variable for the i^{th} unit; $\mathbf{y}_{n_1} = (y_1, \ldots, y_{n_1})$
- $\hat{y}_i$ is the predicted value for an unobserved unit
- $\mathbf{Z}_i$ are covariates (either known for all i or only for units that have provided information)
- $p = p(\mathbf{y}_{n_1}, \mathbf{Z})$, fraction of n_2 units predicted to respond
- Compute,

$$e_1 = \frac{\sum_1^{n_1} y_i + \sum_{n_1+1}^{n} \hat{y}_i}{n}$$

$$e_2 = \frac{\sum_1^{n_1+\mathbf{pn_2}} y_i + \sum_{n_1+\mathbf{pn_2}+1}^{n} \hat{y}_i}{n}$$

[14] Wagner, J. and Raghunathan, T. E. (2010) A new stopping rule for surveys. *Statist. Med.*, 29: 1014–1024.

Extension of Wagner & Raghunathan[14]

- Have information on n_1 of the n units in the sampling frame
- y_i is the observed variable for the i^{th} unit; $\mathbf{y}_{n_1} = (y_1, \ldots, y_{n_1})$
- $\hat{y}_i$ is the predicted value for an unobserved unit
- $\mathbf{Z}_i$ are covariates (either known for all i or only for units that have provided information)
- $p = p(\mathbf{y}_{n_1}, \mathbf{Z})$, fraction of n_2 units predicted to respond
- Compute,

$$e_1 = \frac{\sum_1^{n_1} y_i + \sum_{n_1+1}^{n} \hat{y}_i}{n}$$

$$e_2 = \frac{\sum_1^{n_1+\mathbf{pn_2}} y_i + \sum_{n_1+\mathbf{pn_2}+1}^{n} \hat{y}_i}{n}$$

- Stop data collection when,

$$\text{pr}(|e_1 - e_2| < \delta \mid \text{data, prediction model, } \ldots) > 1 - \gamma$$

- Accommodate stochastic uncertainty: replace $\mathbf{pn_2}$ by a **Binomial** (n_2, p) r.v.
- Additional accommodation:
  Use a **beta-binomial distribution** that injects (posterior) uncertainty in p

[14] Wagner, J. and Raghunathan, T. E. (2010) A new stopping rule for surveys. *Statist. Med.*, 29: 1014–1024.

Using the Binomial(n_2, p) distribution (core idea)

- p = response probability
- n_{goal} = desired number of responses
- γ = probability of obtaining at least n_{goal} responses

- The table presents the required number of contacts to ensure that,

$$\text{pr}(\#\{responses\} \geq n_{goal} \mid p) \geq \gamma$$

	$p = 0.25$		$p = 0.50$		
$\gamma \downarrow$	50	200	50	200	$\leftarrow n_{goal}$
.50	203	803	101	401	
.95	247	887	**119**	436	
Increase over $\gamma = .50$	45	84	18	35	

Using the Binomial(n_2, p) distribution (core idea)

- p = response probability
- n_{goal} = desired number of responses
- γ = probability of obtaining at least n_{goal} responses

- The table presents the required number of contacts to ensure that,

$$\text{pr}(\#\{responses\} \geq n_{goal} \mid p) \geq \gamma$$

	$p = 0.25$		$p = 0.50$		
$\gamma \downarrow$	50	200	50	200	$\leftarrow n_{goal}$
.50	203	803	101	401	
.95	247	887	**119**	436	
Increase over $\gamma = .50$	45	84	18	35	

Beta-binomial Bayes: $p \sim \text{Beta}(\mu, M)$, M is effective sample size

- For $\mu = .50$ and $n_{goal} = 50$ :: $\gamma = .50$, all M: need 101 contacts

$\gamma = 0.95$:

M	5	50	∞
#{required}	230	130	**119**

Timely & Accurate Data are Essential

Data delay matrix from a clinical trial

Currentness of Adverse Events Reporting

AE	Cut-off Date					
	21Oct2018	21Nov2018	21Dec2018	21Jan2019	21Feb2019	21Mar2019
	All / Serious	**All** / Serious	**All** / Serious	**All** / Serious	**All** / Serious	**All** / Serious
Occurred Up To 21Oct2018	**1393** / 163	**1739** / 171	**1781** / 172	**1819** / 172	**1840** / 174	**1848** / 174
Occurred Up To 21Nov2018	**N.A.** / N.A.	**1774** / 180	**1860** / 185	**1900** / 185	**1933** / 187	**1948** / 187
Occurred Up To 21Dec2018	**N.A.** / N.A.	**N.A.** / N.A.	**1870** / 188	**1923** / 189	**1971** / 192	**2004** / 193
Occurred Up To 21Jan2019	**N.A.** / N.A.	**N.A.** / N.A.	**N.A.** / N.A.	**1942** / 195	**2005** / 199	**2051** / 201
Occurred Up To 21Feb2019	**N.A.** / N.A.	**N.A.** / N.A.	**N.A.** / N.A.	**N.A.** / N.A.	**2027** / 204	**2084** / 208
Occurred Up To 21Mar2019	**N.A.** / N.A.	**N.A.** / N.A.	**N.A.** / N.A.	**N.A.** / N.A.	**N.A.** / N.A.	**2093** / 210

A similar approach can (should?) be used in surveys

Care is Needed

- **Validity and efficiency** of data generated by an adaptive design are strongly dependent on protocol-specifics and their alignment with underlying truths
- **Adaptation adds complexity**, requires sophisticated and reliable infrastructure, requires effective training and supervision
- **Valid analyses** of data generated by adaptive methods requires more care and sophistication than those generated from a non-adaptive design
- Consequently, **adaptive designs must be robust** to credible model misspecification and other violations of working assumptions
- **Aggressive simulations** are essential; in the clinical trials context, see[15,16]
- **Research is needed** on the trade-offs between efficiency and robustness, on the policy or clinical consequences of reduced quality, on cost/benefit
- **Stopping rules are important, but so are starting rules**
  - Are the potential benefits of adaptation worth the overhead and risk?

[15] FDA (2016). Adaptive designs for medical device clinical studies: guidance for industry and Food and Drug Administration staff.

[16] FDA (2018). Adaptive designs for clinical trials of drugs and biologics. Guidance for Industry.

Care is Needed

- **Validity and efficiency** of data generated by an adaptive design are strongly dependent on protocol-specifics and their alignment with underlying truths
- **Adaptation adds complexity**, requires sophisticated and reliable infrastructure, requires effective training and supervision
- **Valid analyses** of data generated by adaptive methods requires more care and sophistication than those generated from a non-adaptive design
- Consequently, **adaptive designs must be robust** to credible model misspecification and other violations of working assumptions
- **Aggressive simulations** are essential; in the clinical trials context, see[15,16]
- **Research is needed** on the trade-offs between efficiency and robustness, on the policy or clinical consequences of reduced quality, on cost/benefit
- **Stopping rules are important, but so are starting rules**
  - Are the potential benefits of adaptation worth the overhead and risk?

You get only one chance to generate the data, so don't mess it up

[15] FDA (2016). Adaptive designs for medical device clinical studies: guidance for industry and Food and Drug Administration staff.

[16] FDA (2018). Adaptive designs for clinical trials of drugs and biologics. Guidance for Industry.

#thankyou

Additional Literature

Dworak, P. and Chang, W. (2015) SMART on health and retirement study. American Association for Public Opinion Research A. Conf., Hollywood.

Elsäßer, A., Regnstrom, J., Vetter, T., Koenig, F., Hemmings, R. J., Greco, M., Papaluca-Amati, M. and Posch, M. (2014) Adaptive clinical trial designs for European marketing authorization: a survey of scientific advice letters from the European Medicines Agency. Trials, 15, no. 1, article 383.

European Medicines Agency (2007) Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. Technical Report. Committee for Medicinal Products for Human Use, European Medicines Agency, London.

Hatfield, I., Allison, A., Flight, L., Julious, S. A. and Dimairo, M. (2016) Adaptive designs undertaken in clinical research: a review of registered clinical trials. Trials, 17, article 150.

Jennison, C. and Turnbull, B. W. (1999) Group Sequential Methods with Applications to Clinical Trials. London: Chapman and Hall.

Lin, M., Lee, S., Zhen, B., Scott, J., Horne, A., Solomon, G. and Russek-Cohen, E. (2016) CBERS experience with adaptive design clinical trials. Therp. Innovn Reglatry Sci., 50, 195203.

Lundquist, P. and Särndal, C.-E. (2013) Aspects of responsive design with applications to the Swedish living conditions survey. J. Off. Statist., 29, 557–582.

Mistry, P., Dunn, J. A. and Marshall, A. (2017) A literature review of applied adaptive design methodology within the field of oncology in randomised controlled trials and a proposed extension to the CONSORT guidelines. BMC Med. Res. Methodol., 17, article 108.

Morgan, C. C., Huyck, S., Jenkins, M., Chen, L., Bedding, A., Coffey, C. S., Gaydos, B. and Wathen, J. K. (2014) Adaptive design: results of a 2012 survey on perception and use. Therp. Innovn Reglatry Sci., 48, 473481.

Additional Literature (continued)

Murphy, S. A. (2003) Optimal treatment regimens. J. R. Statist. Soc. B, 65, 331355.

Rao, R. S., Glickman, M. E. and Glynn, R. J. (2008) Stopping rules for surveys with multiple waves of nonrespondent follow-up. Statist. Med., 27, 2196–2213.

Robins, J. M. (2004) Optimal structural nested models for optimal sequential decisions. In Proc. 2nd Seattle Symp. Biostatistics. New York: Springer.

Rosenblum, M., Miller, P., Reist, B., Stuart, E., Thieme, M., and Louis, T. (2019) Adaptive Design in Surveys and Clinical Trials: Similarities, Differences, and Opportunities for Cross-Fertilization. Journal of the Royal Statistical Society, Series A (Statistics in Society). 182, 963-982. https://doi.org/10.1111/rssa.12438

Särndal, C.-E. (2008) Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator. J. Off. Statist., 24, no. 2, article 167.

Särndal, C.-E. (2011) Dealing with survey nonresponse in data collection, in estimation. J. Off. Statist., 27, no. 1, article 1.

Särndal,C.-E. and Lundström, S. (2010) Design for estimation: identifying auxiliary vectors to reduce nonresponse bias. Surv. Methodol., 36, 131144.

Scharfstein, D. O., Tsiatis, A. A. and Robins, J. M. (1997) Semiparametric efficiency and its implication on the design and analysis of group-sequential studies. J. Am. Statist. Ass., 92, 13421350.

Schouten, B., Cobben, F. and Bethlehem, J. (2009) Indicators for the representativeness of survey response. Surv. Methodol., 35, 101113. Schouten, B., Shlomo, N. and Skinner, C. (2011) Indicators for monitoring and improving survey response. J. Off. Statist., 27, 231253.

van der Laan, M. J. and Luedtke, A. R. (2015) Targeted learning of the mean outcome under an optimal dynamic treatment rule. J. Causl Inf., 3, 6195.

Representativeness: Xiao-Li Meng's Cautionary Tale[17,18]

(A big sample size, n, may not save the day)

- Compare the MSE for two estimators of the finite population mean ($\bar{Y}_N$), N large

  $\bar{y}_{srs}$: Sample mean of a simple random sample of size $n_{srs} = 100$
  $\bar{y}_{sel}$: A self-selected, web sample of size n_{sel}

- With $\rho(\mathbf{Y}, \boldsymbol{\pi}) = \text{cor}(\mathbf{Y}, \text{inclusion propensity}) = 0.05$, and $frac = n_{sel}/N$,

$$\text{MSE}_{sel} \leq \text{MSE}_{srs} \iff frac \geq 20\%$$

- For example, $N = 50M$ requires $n_{sel} \geq 10M$ to beat the SRS with $n_{srs} = 100$ (!)
- Good information on $\rho(\mathbf{Y}, \boldsymbol{\pi})$ is needed to rescue the situation

A large sampling fraction, n/N, may not be protective

[17] Meng's discussion of Keiding&Louis (2016)

[18] Meng (2018). Statistical Paradises and Paradoxes in Big Data (I): Law of Large Populations, Big Data Paradox, and the 2016 Presidential Election. *Annals of Applied Statistics*, 12: 685–726.