Non-Traditional RCTs
Standing Committee on Evidence Synthesis, and Communications in Diet and Chronic Disease Relationships

Robert M. Califf MD
Head of Clinical Policy and Strategy, Verily Life Sciences and Google Health
Adjunct Professor, Duke and Stanford
August 31st, 2021

Stating the Obvious

- People want to know: “What should I eat to stay healthy or treat my problem?”
- Diet is not a simple intervention
  - Behavior over time
  - Dosing difficult or impossible to quantify
  - Complex constituents of intervention with complex interactions with human biology
- Traditional RCTs seem ill-suited for the task of sorting out this complexity

Overview

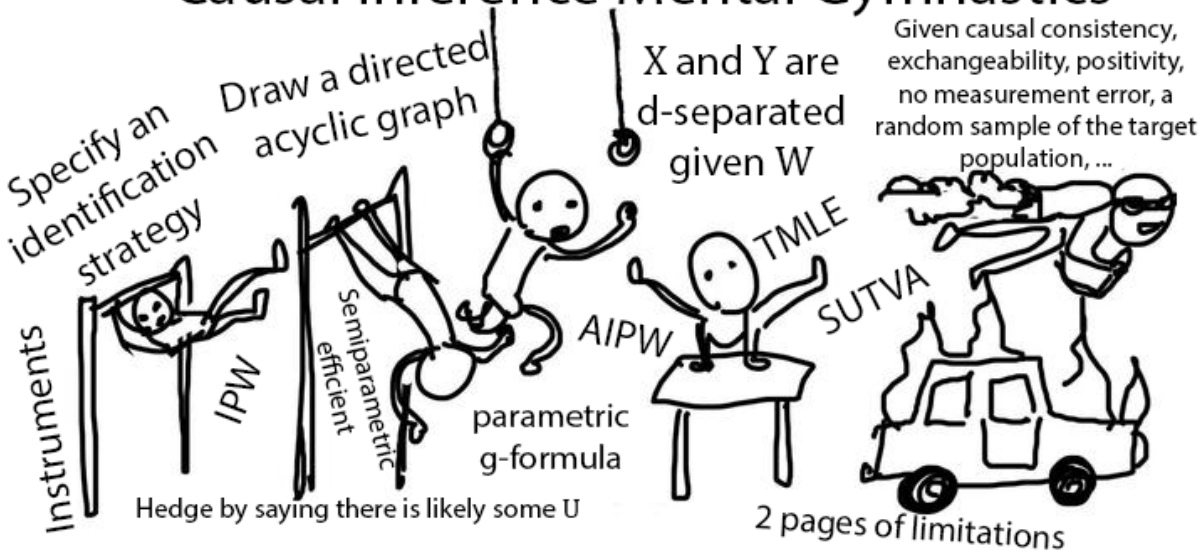
- Why do we need randomization to have confidence when we intend to make a causal inference?
- What do we mean by ”traditional” and “non-traditional” RCTs?
- What are key types of ”non-traditional” designs

Why do we Need to Randomize?

- Eliminates bias in treatment assignment
- Assures that both measured and unmeasured confounders are randomly distributed in intervention groups
- Sets a true “time zero” or inception point to assess the outcomes as an “intention to treat”
- Permits the use of probability theory to express the likelihood that differences in outcome according treatment groups are likely to be due to chance

Risk Factor Mental Gymnastics

Causal Inference Mental Gymnastics



Does the design emulate a hypothetical randomized trial design?	• Designing nonrandomized studies to emulate a hypothetical RCT can help define the study inception point and avoid major design flaws that contribute to bias.
Is the comparator or control condition appropriate?	• Active comparator groups composed of patients using treatments with similar indications and modality as the treatment of interest can greatly mitigate the risk of unmeasured confounding. • Non-user comparator groups are suspect, as patients who are not receiving treatment are often very different.
Does the primary analysis adjust for measured confounders?	• Propensity score matching or weighting allows for evaluation of balance on measured confounders, similar to the evaluation of balance in an RCT. • Optimal adjustment includes all risk factors for outcome. • Automated variable selection methods can aid investigator judgement in selection of adjustment variables.
Do sensitivity analyses quantify the potential impact of residual confounding?	• Evaluate potential residual confounding through assessment of balance on confounders measured only in a subset, assessment of control outcomes, quantitative bias analysis, or instrumental variable methods.
Are methods open to inspection and (if possible) replication?	• Registration of study protocols prior to conducting analyses of treatment effects promotes transparency and can mitigate concerns about data dredging. • Supporting replication of study findings in a separate data source, providing access to de-identified data sets, and sharing analytic code all contribute to confidence in study findings.

J Franklin et al; Clin Pharmacol Ther. 2021 Apr 7.
doi: 10.1002/cpt.2255

SOUNDING BOARD FREE PREVIEW

The Magic of Randomization versus the Myth of Real-World Evidence

Rory Collins, F.R.S., Louise Bowman, M.D., F.R.C.P., Martin Landray, Ph.D., F.R.C.P., and Richard Peto, F.R.S.

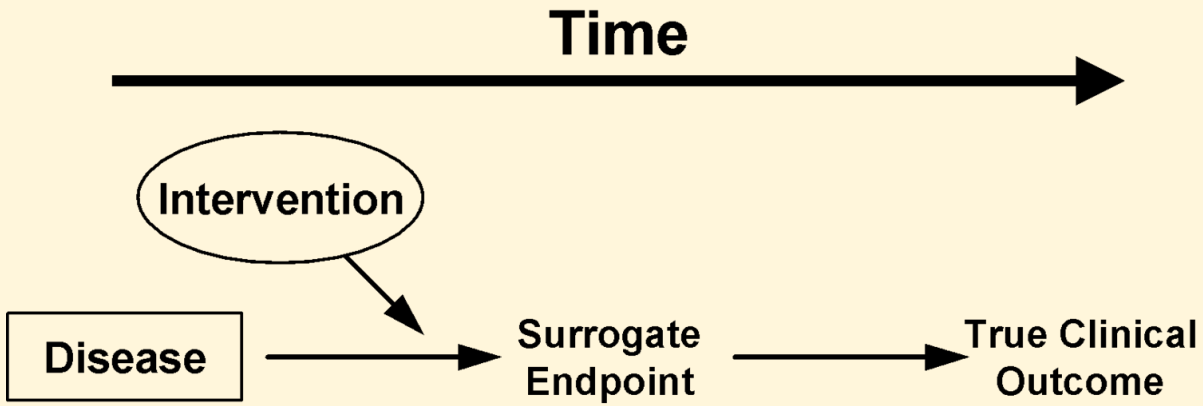
Nonrandomized observational analyses have been promoted as alternatives to randomized clinical trials. However, randomization ensures balance between groups, whereas nonrandomized studies are often biased by between-group differences. Efforts to reduce the cost and complexity of clinical trials are preferable to relying on observational studies.

February 13, 2020
N Engl J Med 2020; 382:674-678
DOI: 10.1056/NEJMsb1901642

Related Articles

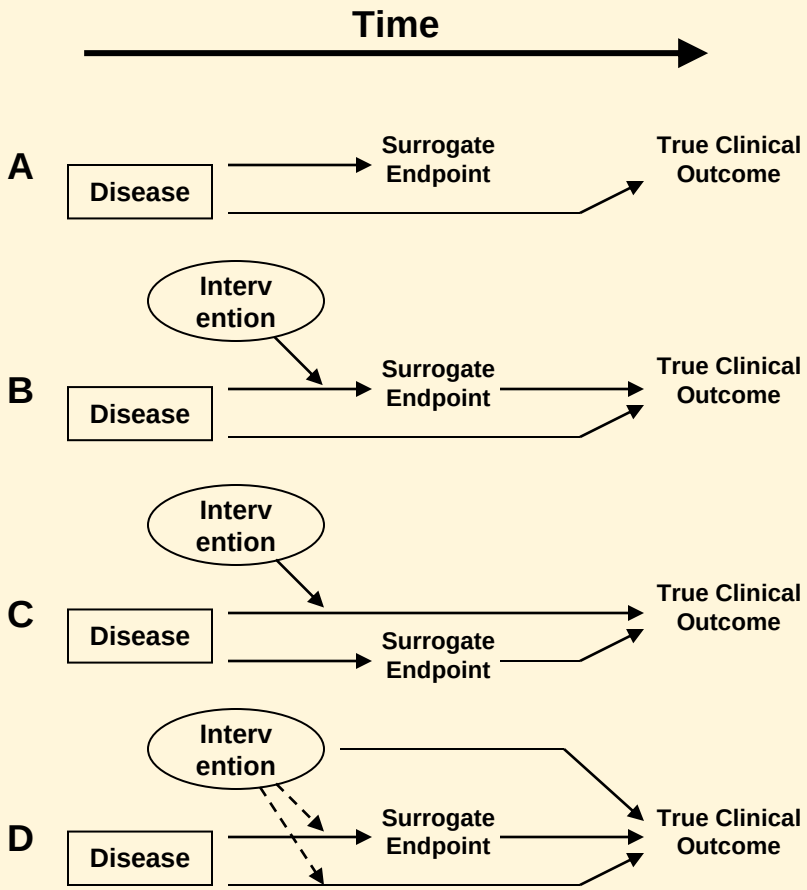
But randomization is just one tool (albeit an essential one) in the chest to understand the effect of dietary recommendations or actual diet eaten

"Even in the best of circumstances, it is possible for surrogate endpoints to be misleading by either overestimating or underestimating an intervention's effect on clinical outcomes."



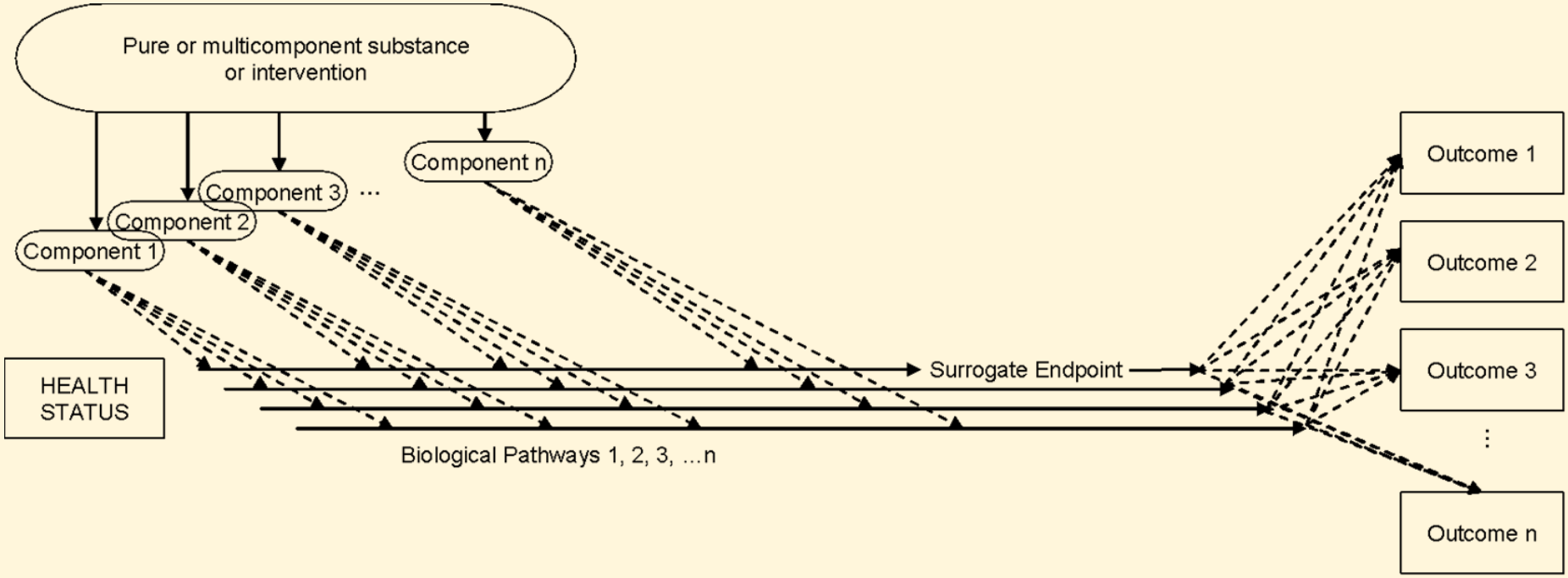
Fleming, T. R., and D. L. DeMets. 1996. Surrogate end points in clinical trials: Are we being misled? *Annals of Internal Medicine* 125(7):605–613.

Failures of Surrogate Endpoints



Fleming, T. R., and D. L. DeMets. 1996. Surrogate end points in clinical trials: Are we being misled? *Annals of Internal Medicine* 125(7):605–613.

Biological Complexity Leads to Many Opportunities for Error



What is a “Traditional RCT”?

- Simple randomization
- Rigid inclusion/exclusion criteria
- Highly regimented intervention
- Blinded
- Measurement of many well-defined baseline characteristics and outcomes
- Much attention to bureaucratic ”quality”
- Expensive-–time consuming
- Conducted in specialized, labor-intensive environments
- Primary goal is validity of the experiment

Purposes of Trials

- Explanatory
  - What is the mechanism by which an intervention causes an effect
- Pragmatic
  - Does an intervention cause an effect that informs decision makers (people, patients, carers, clinicians, health systems, policy makers) about their next decision?
  - Quality by design is essential
    - https://ctti-clinicaltrials.org/category/topics/quality/quality-by-design/

Important Meta-Design Variations

- Adaptive designs
- Factorial designs
- Master protocols
  - Umbrella
  - Basket
  - Platform
- N of 1 trials

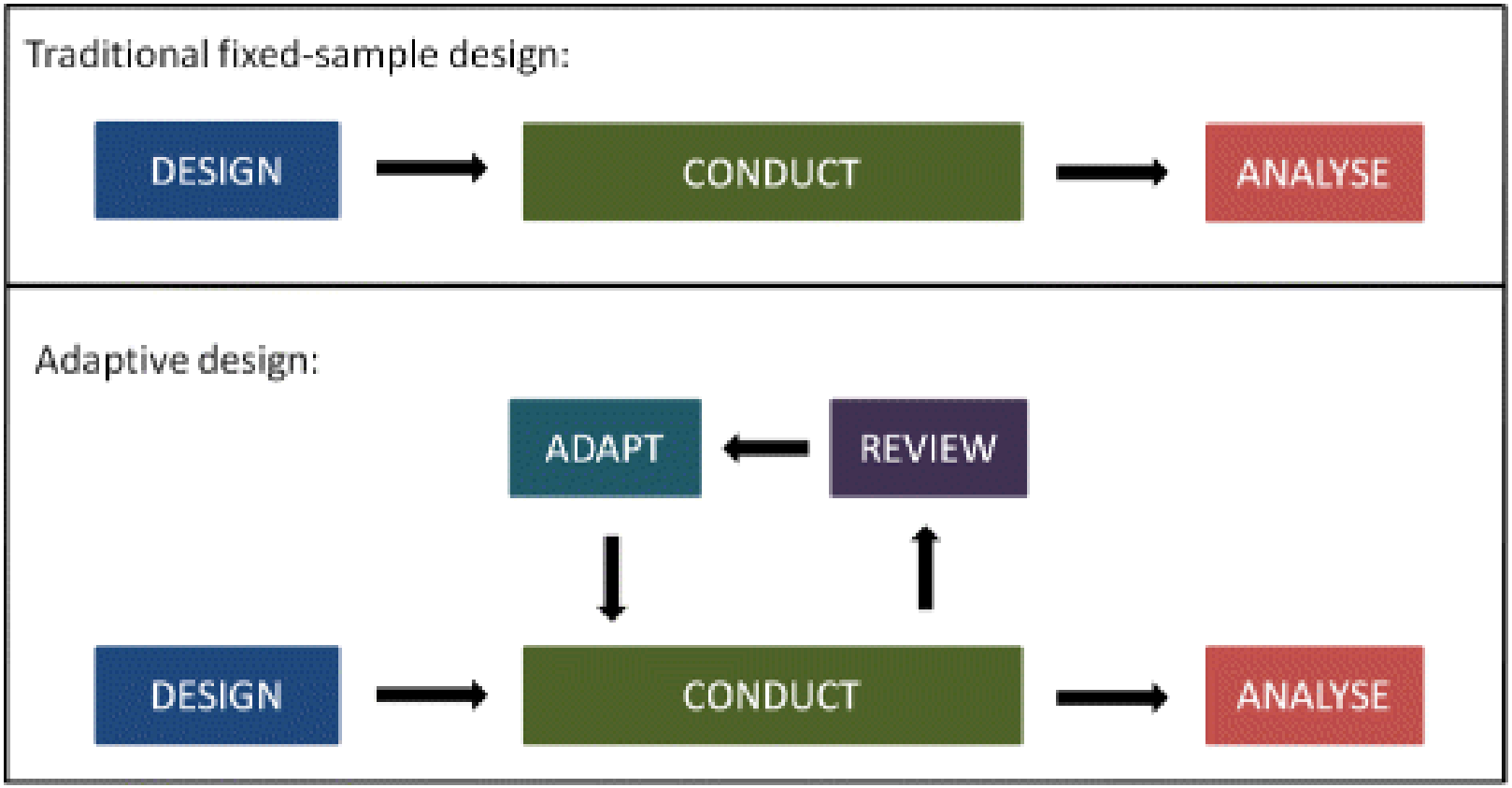
Therapeutic Principles

- Treatment effects usually modest
- Qualitative interactions are uncommon
- Quantitative interactions common
- Unintended targets common
- Long-term vs. short-term effects may differ
- Combinations are unpredictable
- Class effect may not be valid
- Most treatments produce a mixture of benefits and risks

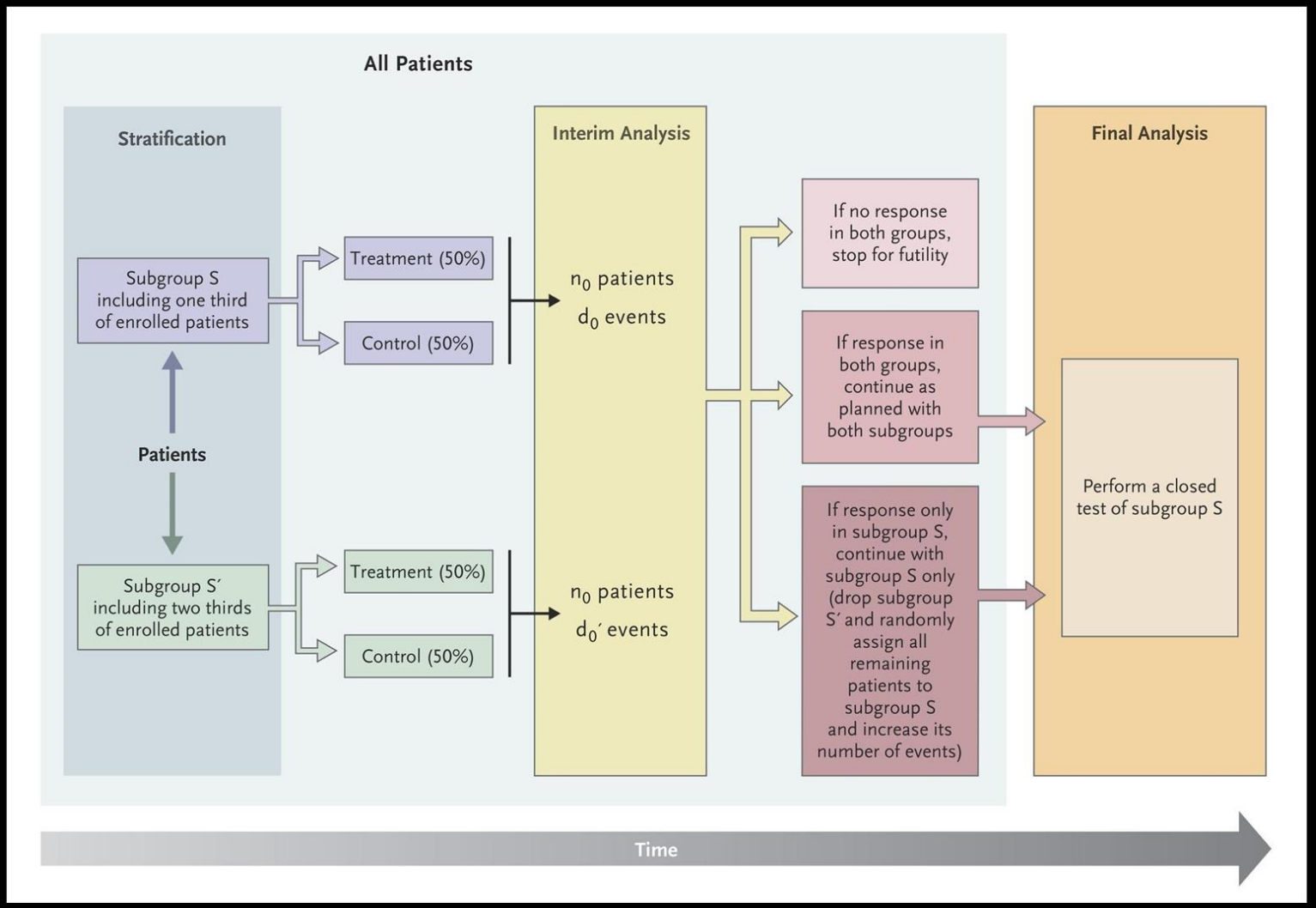
Califf and DeMets Circulation 2002: 1015-21

Ways to Randomize

- Simple
- Blocked
- Adaptive
  - Covariate adaptive
  - Response adaptive


Traditional fixed-sample design:
DESIGN
CONDUCT
ANALYSE
Adaptive design:
ADAPT
REVIEW
DESIGN
CONDUCT
ANALYSE

Schematic Representation of an Adaptive Two-Stage Population-Enrichment Design.



Bhatt DL, Mehta C. N Engl J Med 2016;375:65-74

Types of Adaptive Designs.

Table 1. Types of Adaptive Designs.*

Stage of Development and Design Type	Strength	Weakness
First-in-human, phase 1 design with goal of establishing the MTD		
Single ascending dose or multiple ascending doses	Establishes MTD and biologic activity	Uses larger cohort sizes at potentially safe doses
Dose-escalation, 3+3, continual reassessment method, Bayesian logistic-regression method, or modified toxicity probability interval design	Provides more accurate estimate of MTD with smaller cohort size	May yield more cases of toxic events than design with single ascending dose or multiple ascending doses
Phase 2 design, with the goal of establishing efficacy and choosing doses for phase 3 trial		
Fixed-sample design, traditional proof-of-concept design (MTD vs. placebo), or dose-ranging design (MTD, placebo, and intermediate doses)	Is simple to implement and easy to design	Has less precision than adaptive design
Adaptive design with proof of concept, early stopping and sample-size reestimation, dose ranging with selection at interim analysis, or dose ranging with frequent Bayesian adaptation of randomization ratios	Yields more precise estimates for same sample size	Is more complex to implement; requires more lead time to set up; operating characteristics determined only by simulation
Seamless phase 2–3 design		
Operationally seamless	Eliminates time between phase 2 and phase 3; permits sponsor involvement for dose selection at the end of phase 2; uses conventional final analysis, parameter estimates, and confidence intervals	Has a final analysis based only on data from phase 3
Inferentially seamless	Combines data from both phases for final analysis	Has no sponsor involvement in dose selection at end of phase 2; has risk of inadequate dose–response modeling; has a complex final analysis involving closed testing; uses nonconventional parameter estimates and confidence intervals; data from the two phases may not be homogeneous
Sample-size reestimation		
With blinded data	Uses conventional final analysis; has fewer regulatory hurdles	Allows sample-size adjustments due to unknown variance only
With unblinded data	Allows sample-size adjustments due to unknown treatment effect or unknown variance; can determine sample size after review of actual data from the trial instead of from pilot studies	Interim estimate of treatment effect can be misleading; requires strict firewalls to prevent leakage of information about adaptive rules or decisions; potential for operational bias if investigator behavior changes; requires meticulous up-front planning; uses nonconventional final analysis with prespecified weighting of the cohorts before and after sample-size reestimation; may face regulatory hurdle
Group sequential design		
Classic	Enables early stopping for efficacy, futility, or harm; has flexible alpha spending functions; can alter maximum sample size in a blinded manner	Cannot alter maximum sample size or events in an unblinded manner; uses nonconventional parameter estimates and confidence intervals; if trial terminates early for efficacy, overruns pose risk of downturn from significance to nonsignificance; greater burden on data and safety monitoring committee to review totality of evidence before premature termination†
Adaptive	Includes all advantages of classic group sequential designs; can alter maximum sample size in an unblinded manner; can switch end point from noninferiority to superiority; can alter number and spacing of interim analyses, and alpha spending function, on the basis of unblinded interim analysis; overruns are not a problem since trial proceeds to completion with increased enrollment and resolution of responses in all patients instead of being terminated early with risk of downturn from unknown or unadjudicated responses	Includes all the disadvantages of sample-size reestimation with unblinded data; uses nonconventional parameter estimates and confidence intervals
Population-enrichment design	Can eliminate nonperforming subgroups at interim analysis if treatment is effective in selected subgroups only	Must prospectively identify which subgroups to target; may eliminate subgroups in which treatment is effective; loses power as number of targeted subgroups increases; loses power if there is low prevalence of effective subgroups; biomarker cutoff points for subgroup partitioning not known

* MTD denotes maximum tolerated dose.
† In an overrun situation, patients for whom the primary end points are unknown (because of unadjudicated data or delayed response) at the time of an early-termination decision will be included in the final analysis.

Bhatt DL, Mehta C. N Engl J Med 2016;375:65-74

The NEW ENGLAND JOURNAL of MEDICINE

Adaptive Designs for Clinical Trials of Drugs and Biologics

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

November 2019
Biostatistics

Units of Randomization

- Individual
- Cluster
  - Randomize at the group level
  - Avoids contamination of intervention in control group
  - Good for “bundled intervention”
- Stepped wedge design
  - Randomize timing of intervention at the group level
  - Useful to deal with perceptions that getting the intervention will be advantageous, regardless of condition of equipoise based on available data

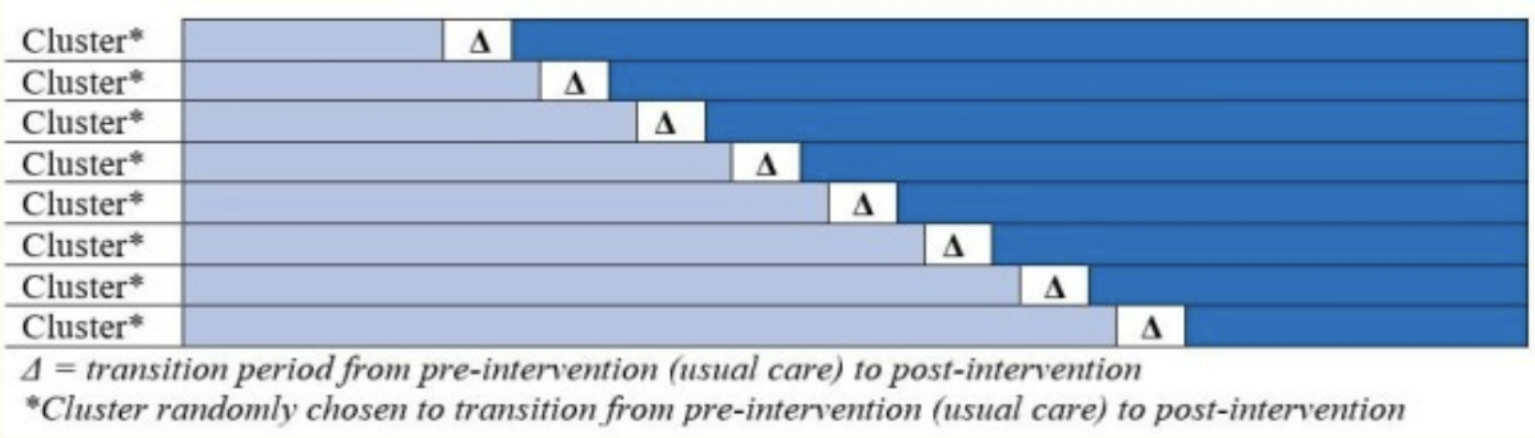


Figure 1

Schematic showing the transition of clusters from usual care to intervention. In a cluster randomized stepped wedge study, all clusters begin the trial providing usual care (preintervention). After a predetermined amount of time in the preintervention period, clusters begin randomly transitioning to the intervention arm. Each cluster provides observations for both the control group (usual care) and the intervention group.

verily

Harvin, JA: Trauma Surg Acute Care Open. 2020; 5(1):e000420.

Types of Master Protocols.

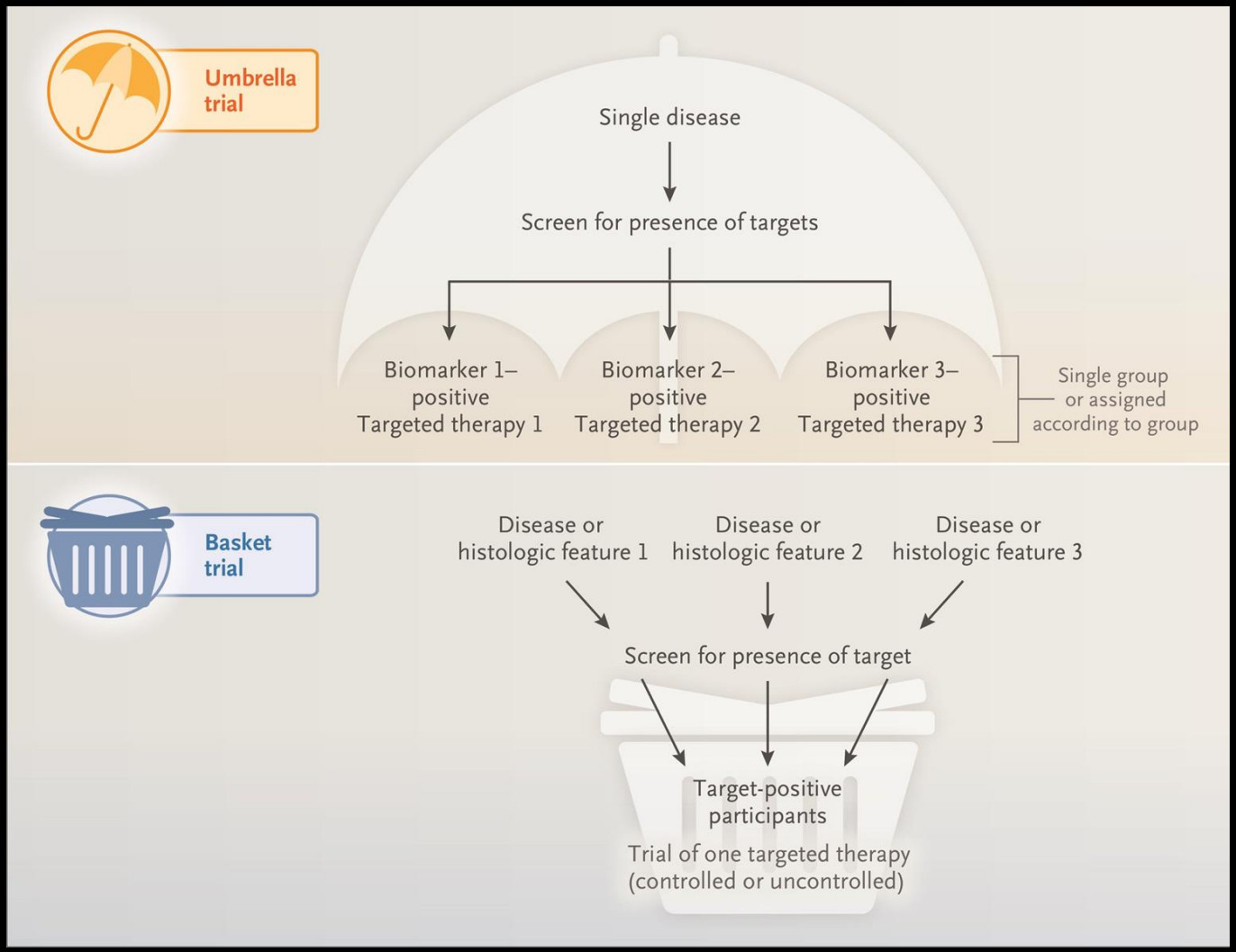
Table 1. Types of Master Protocols.

Type of Trial	Objective
Umbrella	To study multiple targeted therapies in the context of a single disease
Basket	To study a single targeted therapy in the context of multiple diseases or disease subtypes
Platform	To study multiple targeted therapies in the context of a single disease in a perpetual manner, with therapies allowed to enter or leave the platform on the basis of a decision algorithm

Woodcock J, LaVange LM. N Engl J Med ;377:62-70

The NEW ENGLAND JOURNAL of MEDICINE

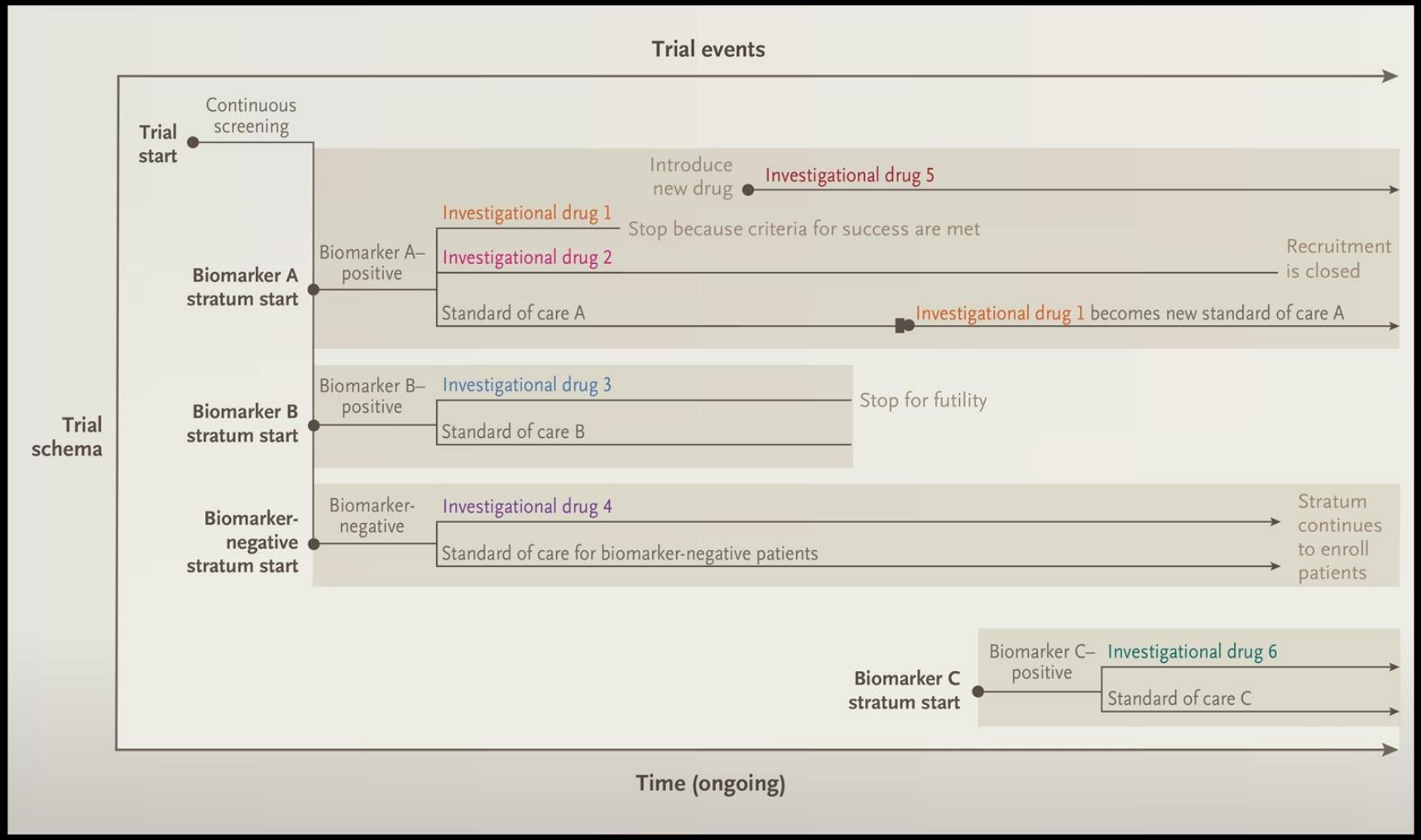
Umbrella Trial and Basket Trial.



Woodcock J, LaVange LM. N Engl J Med ;377:62-70

The NEW ENGLAND JOURNAL of MEDICINE

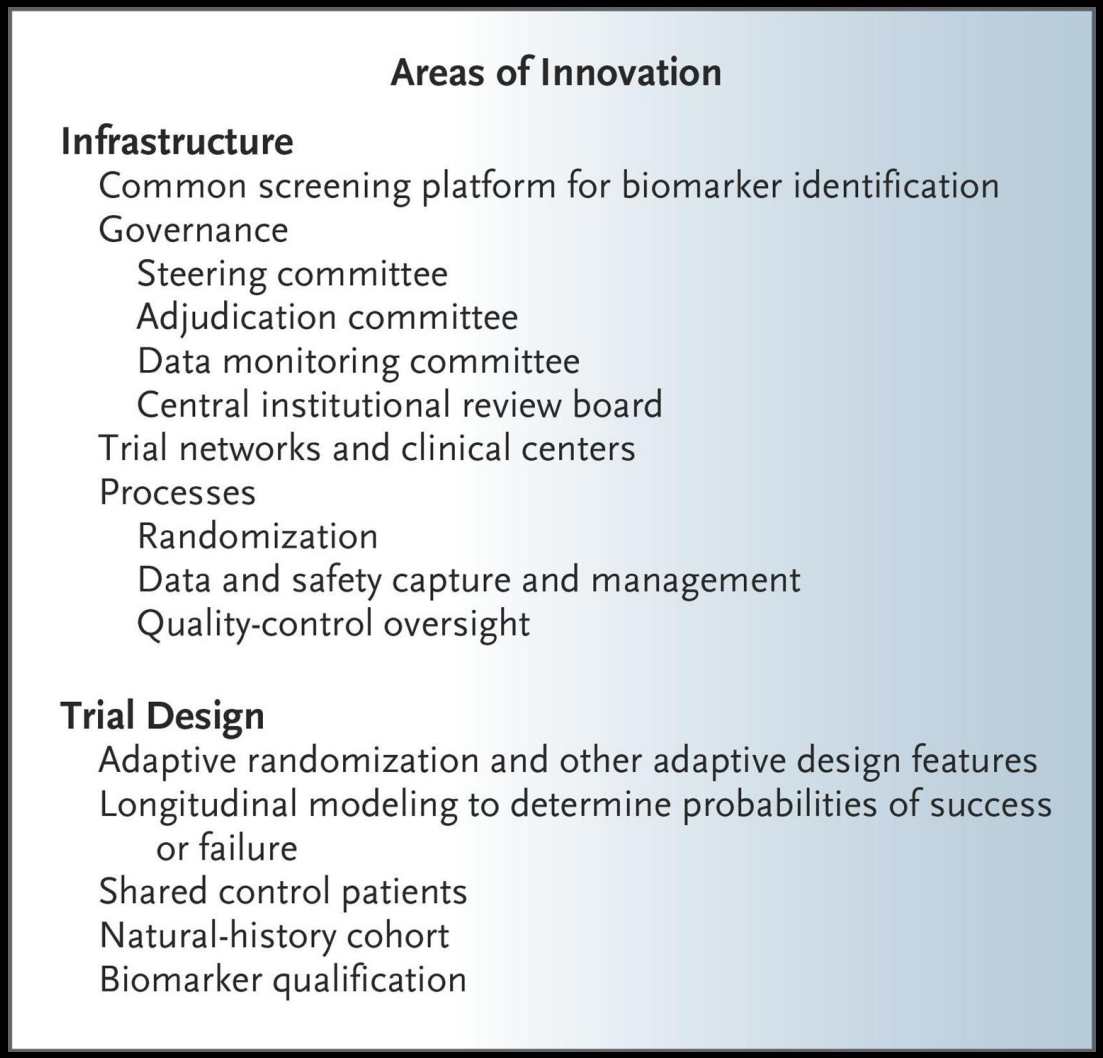
Potential Design of a Platform Trial Involving a Single Disease.



Woodcock J, LaVange LM. N Engl J Med ;377:62-70

The NEW ENGLAND JOURNAL of MEDICINE

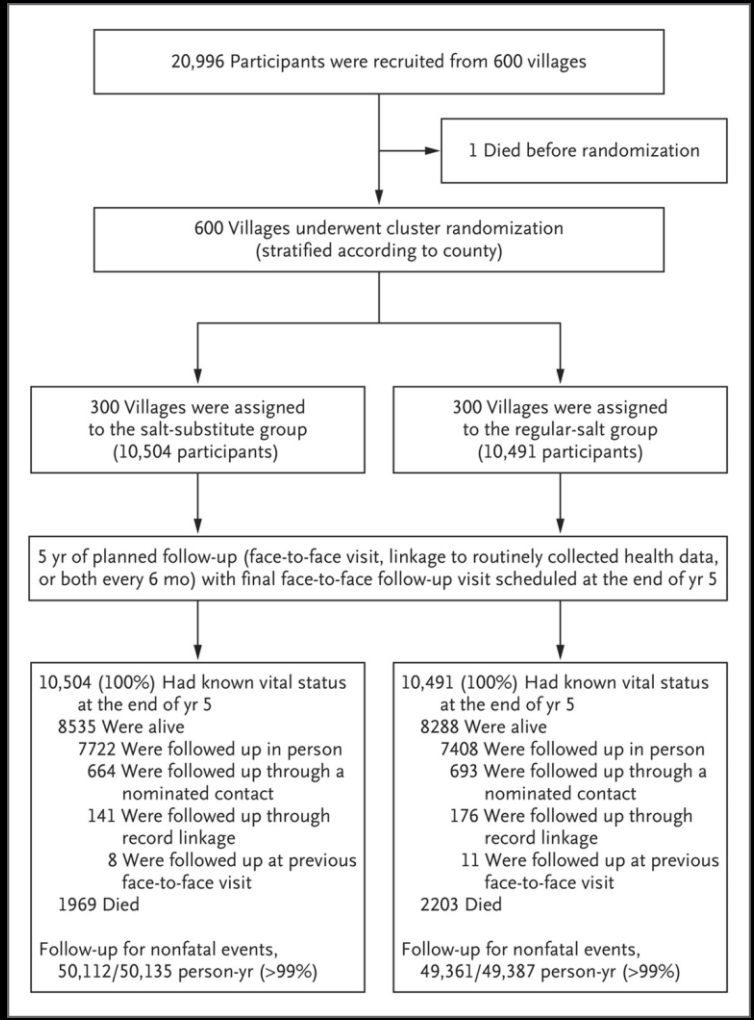
Areas of Innovation in Master Protocols.



Woodcock J, LaVange LM. N Engl J Med ;377:62-70

The NEW ENGLAND JOURNAL of MEDICINE

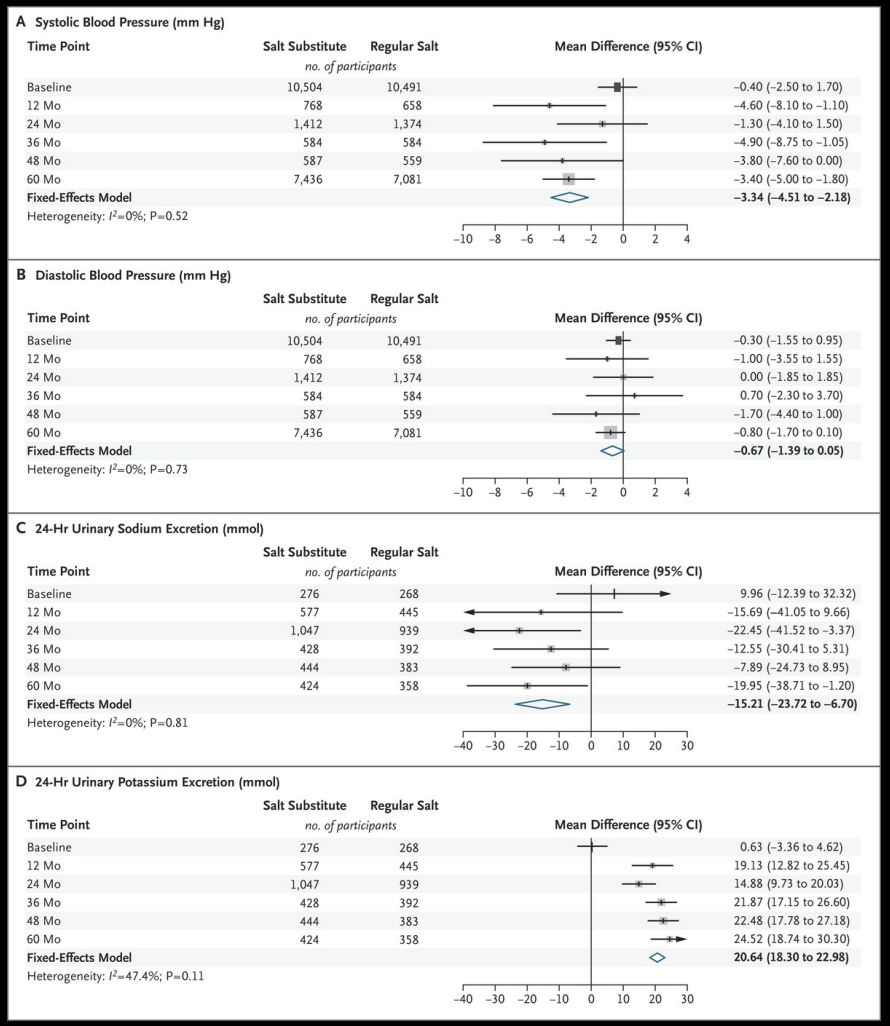
Enrollment, Randomization, and Follow-up.



B Neal et al. N Engl J Med 2021. DOI: 10.1056/NEJMoa2105675

The NEW ENGLAND JOURNAL of MEDICINE

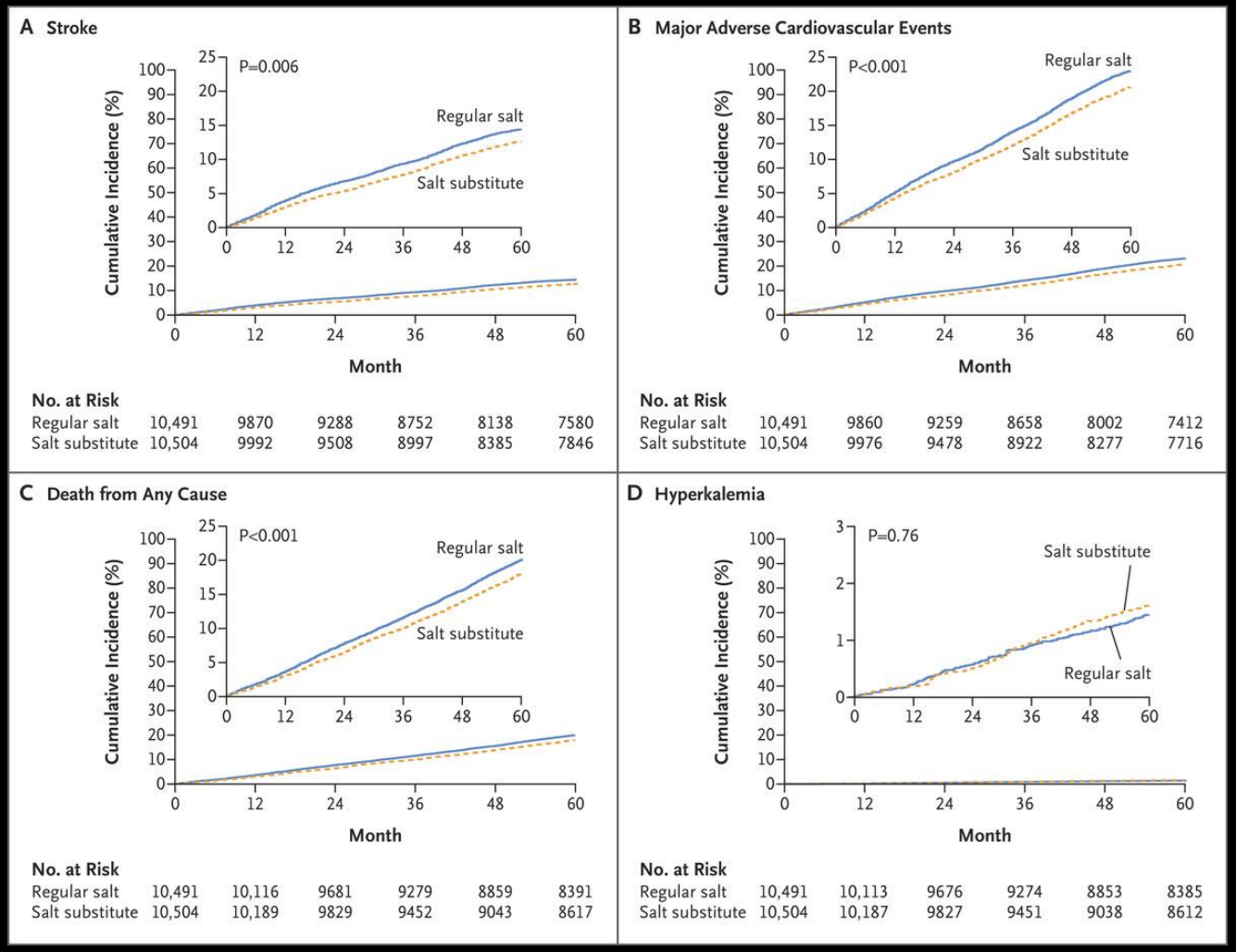
Effects of Salt Substitution on Blood Pressure and 24-Hour Urinary Sodium and Potassium Excretion.



B Neal et al. N Engl J Med 2021. DOI: 10.1056/NEJMoa2105675

The NEW ENGLAND JOURNAL of MEDICINE

Effects of Salt Substitution on Trial Outcomes.

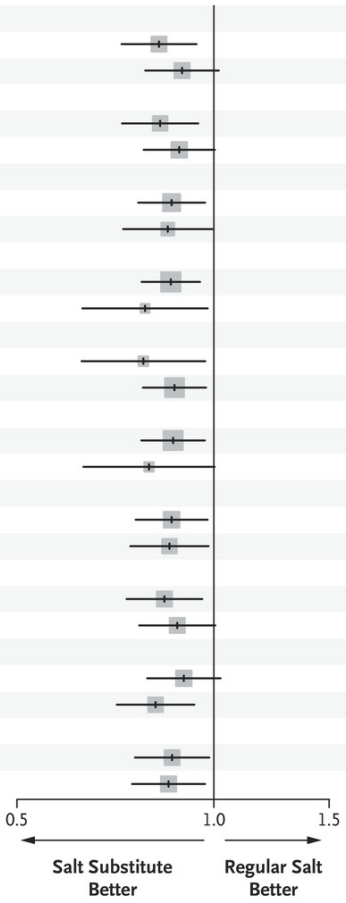


B Neal et al. N Engl J Med 2021. DOI: 10.1056/NEJMoa2105675

The NEW ENGLAND JOURNAL of MEDICINE

Subgroup Analysis of the Effects of Salt Substitution on the Primary Outcome of Stroke.

Subgroup	Salt Substitute (N=10,504)	Regular Salt (N=10,491)	Rate Ratio (95% CI)
	no. of events per 1000 person-yr		
Age			
>65 yr	29.20	35.45	0.82 (0.72–0.94)
≤65 yr	29.09	32.04	0.89 (0.79–1.02)
Sex			
Female	25.14	30.12	0.83 (0.72–0.95)
Male	33.24	37.22	0.89 (0.78–1.00)
Education			
Less educated	28.37	32.67	0.86 (0.77–0.97)
More educated	31.12	36.21	0.85 (0.73–0.99)
Cerebrovascular disease			
Yes	35.97	41.49	0.86 (0.78–0.95)
No	11.75	15.14	0.78 (0.63–0.98)
Diabetes			
Yes	34.34	43.82	0.78 (0.63–0.97)
No	28.56	32.49	0.87 (0.78–0.97)
Hypertensive			
Yes	29.05	33.21	0.87 (0.77–0.97)
No	29.97	37.15	0.80 (0.63–1.00)
Antihypertensive use			
Yes	26.28	30.38	0.86 (0.76–0.98)
No	33.41	38.51	0.86 (0.75–0.98)
Systolic blood pressure			
>153 mm Hg	28.72	34.45	0.84 (0.74–0.96)
≤153 mm Hg	29.54	32.86	0.88 (0.77–1.00)
Diastolic blood pressure			
>89 mm Hg	32.71	36.89	0.90 (0.79–1.02)
≤89 mm Hg	25.64	30.47	0.81 (0.71–0.93)
Body-mass index			
>24.6	27.77	31.56	0.86 (0.76–0.98)
≤24.6	30.53	35.88	0.85 (0.75–0.97)



B Neal et al. N Engl J Med 2021. DOI: 10.1056/NEJMoa2105675

The NEW ENGLAND JOURNAL of MEDICINE

HiLo: A Pragmatic, Randomized Trial of Phosphate Management for Patients on Maintenance Hemodialysis

Setting & Participants

HiLo

Pragmatic, cluster-randomized trial

4,400 patients receiving thrice-weekly hemodialysis in 80-120 dialysis facilities

Intervention

'Hi' phosphate target (≥6.5 mg/dl)

vs

'Lo' phosphate target (<5.5 mg/dl)

Follow-up: 27-45 months

–

Interventions to reach phosphate targets at the discretion of the dietitians & providers

Novel Design Features

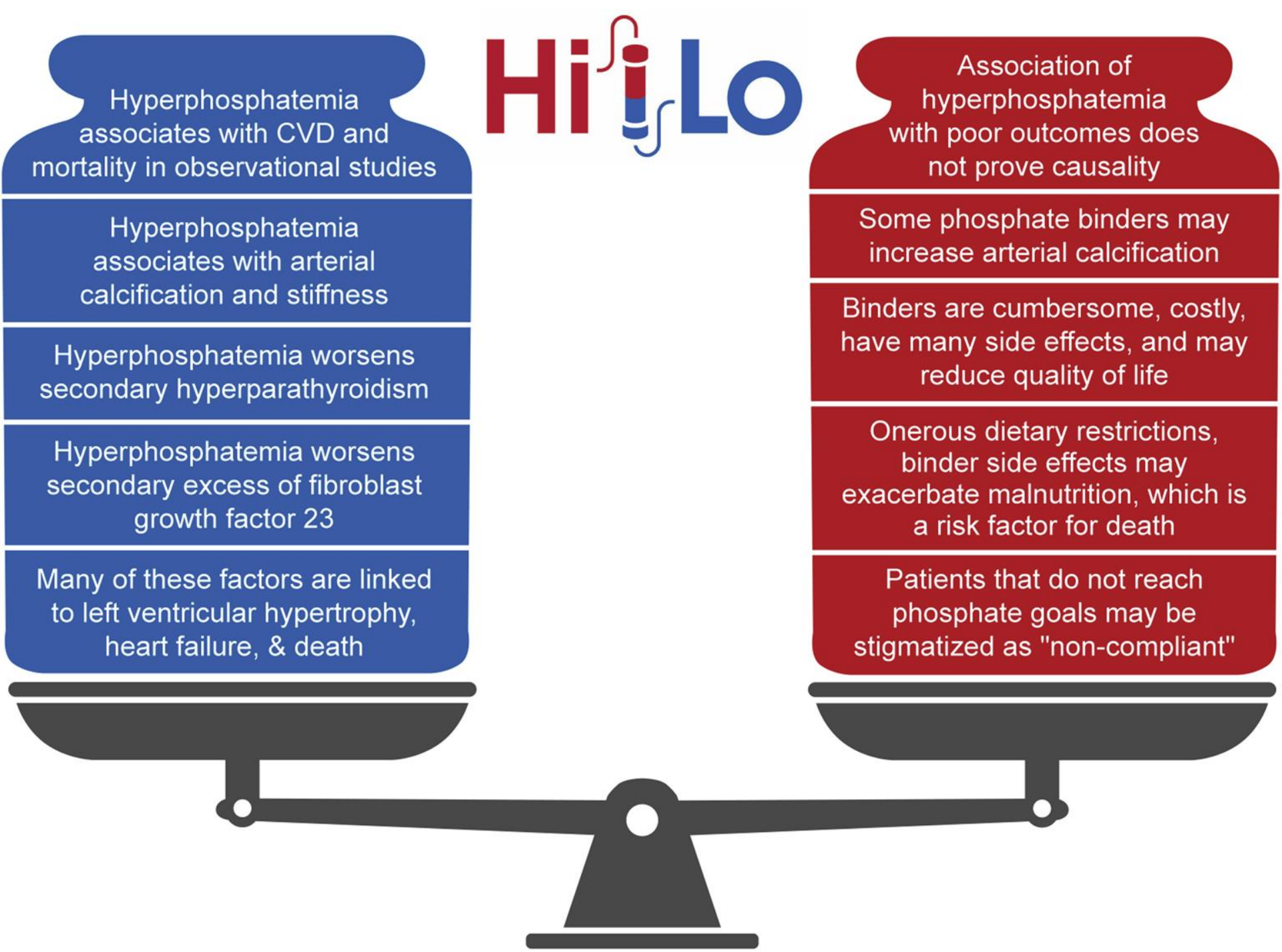
- Extensive stakeholder engagement with patients, dietitians, nephrologists
- Hierarchical composite outcome of all-cause mortality & hospitalizations
- Pragmatic trial with liberal eligibility criteria
- Electronic informed consent (eConsent)
- Real-world data collection from EHR

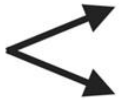
CONCLUSION: HiLo will address the question of what serum phosphate target to use in hemodialysis while advancing methods for pragmatic clinical trials in nephrology.

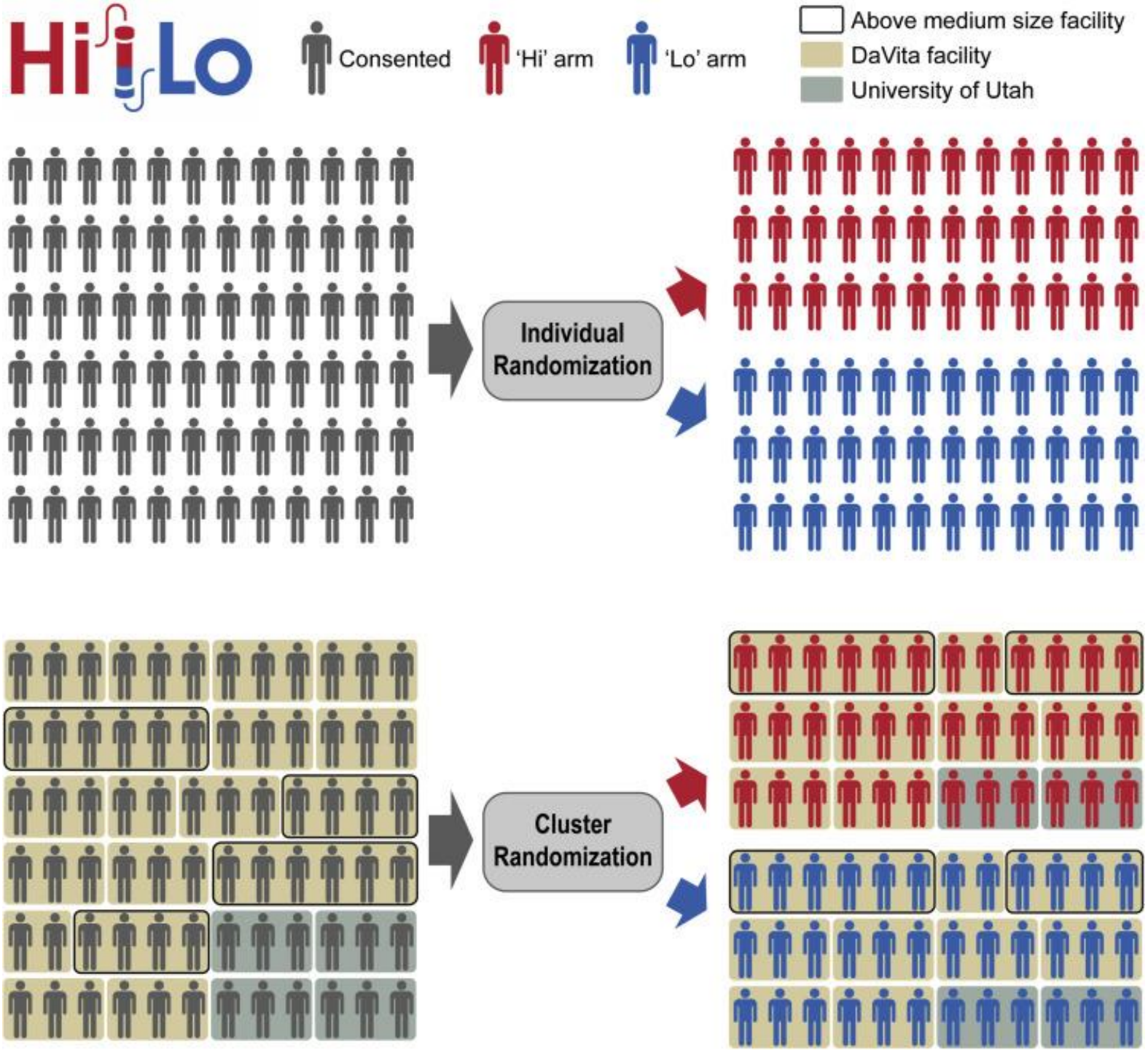
Daniel L. Edmonston, Tamara Isakova, Laura M. Dember, et al (2020)

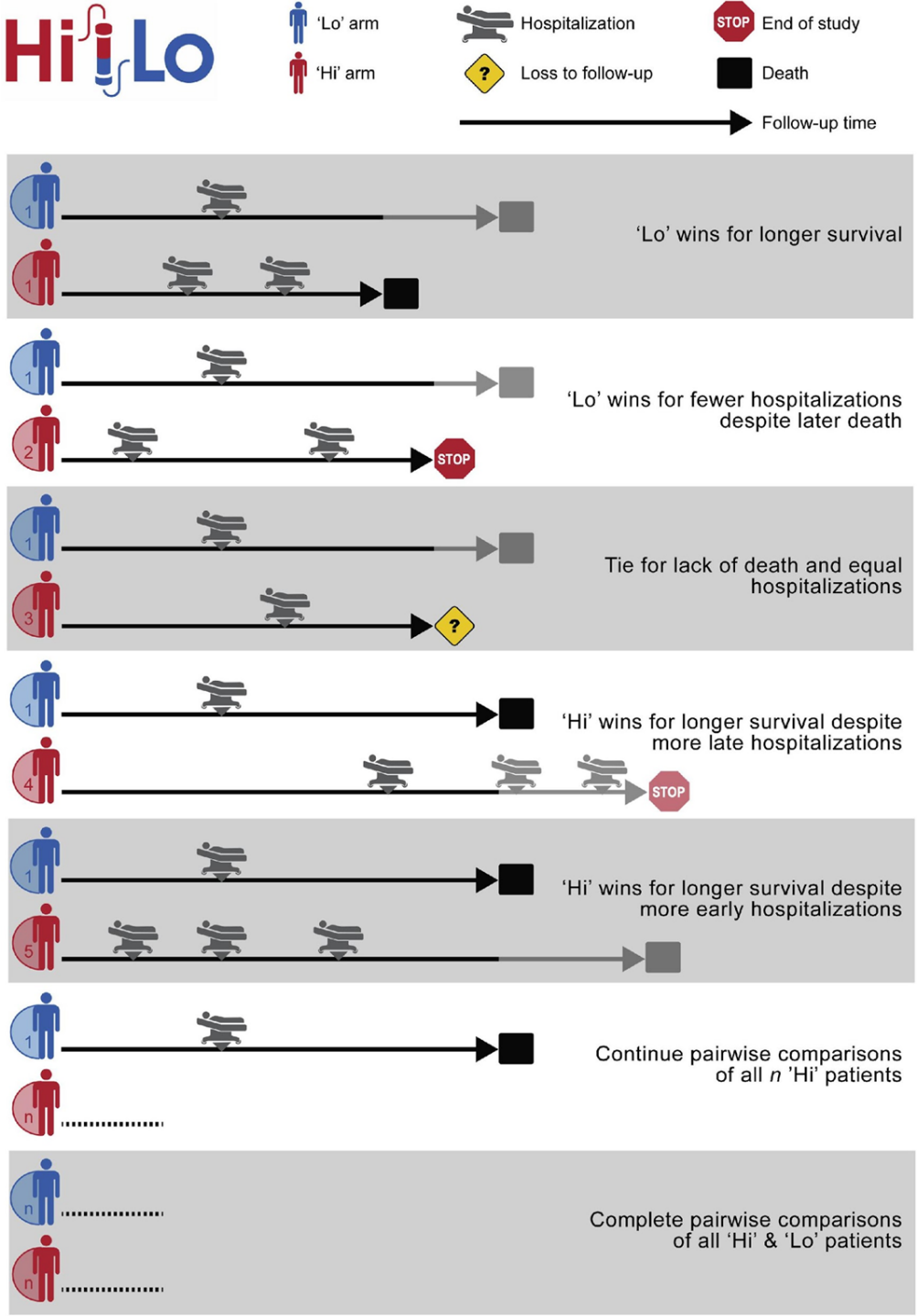
@AJKDonline | DOI: 10.1053/j.ajkd.2020.10.008

AJKD
AMERICAN JOURNAL OF KIDNEY DISEASES


Hi
Lo
Hyperphosphatemia associates with CVD and mortality in observational studies
Hyperphosphatemia associates with arterial calcification and stiffness
Hyperphosphatemia worsens secondary hyperparathyroidism
Hyperphosphatemia worsens secondary excess of fibroblast growth factor 23
Many of these factors are linked to left ventricular hypertrophy, heart failure, & death
Association of hyperphosphatemia with poor outcomes does not prove causality
Some phosphate binders may increase arterial calcification
Binders are cumbersome, costly, have many side effects, and may reduce quality of life
Onerous dietary restrictions, binder side effects may exacerbate malnutrition, which is a risk factor for death
Patients that do not reach phosphate goals may be stigmatized as "non-compliant"

Explanatory Trial		Pragmatic Trial
Strict eligibility criteria based on prior phosphate control		Liberal eligibility criteria irrespective of prior phosphate control
Individual randomization		Cluster randomization
Dedicated study visits outside usual dialysis		Study activities occur during usual dialysis care
Protocolized phosphate interventions led by site investigators		How to reach phosphate targets at discretion of clinical team
Onsite study staff and monitors		No onsite study staff, remote monitors
Informed consent obtained by local study staff		eConsent obtained by central study leadership
Trial-specific data collection via case report forms		Real-world data collection via EHR
Endpoints that require adjudication		Endpoints extracted from EHR without adjudication
Formal adverse event reporting		No formal adverse event reporting
High cost		Lower cost
Extrapolation required for patients that would not meet strict eligibility criteria		Maximize generalizability to US standard in-center hemodialysis population


Hi Lo
Consented
'Hi' arm
'Lo' arm
Above medium size facility
DaVita facility
University of Utah
Individual Randomization
Cluster Randomization


Hi Lo
'Lo' arm
'Hi' arm
Hospitalization
Loss to follow-up
STOP
End of study
Death
Follow-up time
'Lo' wins for longer survival
'Lo' wins for fewer hospitalizations despite later death
Tie for lack of death and equal hospitalizations
'Hi' wins for longer survival despite more late hospitalizations
'Hi' wins for longer survival despite more early hospitalizations
Continue pairwise comparisons of all n 'Hi' patients
Complete pairwise comparisons of all 'Hi' & 'Lo' patients

Model Assumptions

Follow-up time of 2-4 years to simulate time to administrative censoring

Annual loss to follow-up rate of 5%

Annual mortality rates of 15% ('Hi' arm) and 12.8% ('Lo' arm); equivalent to a clinically relevant mortality HR of 0.85

Annual hospitalization rate of 2.0 ('Hi' arm) and 1.89 ('Lo' arm) per patient; equivalent to a conservative 5.5% difference

35% of study population without a single hospitalization

Assume an ICC of 0.001 for mortality and 0.003 for hospitalization

Simulation Results

Simulation 1: 80 clusters with 55 patients each for 5000 iterations

- Mixed Model **86.3% Power**
- Wilcoxon Rank Sum **87.0% Power**

Simulation 2: 120 clusters with 36 patients each for 5000 iterations

- Mixed Model **85.0% Power**
- Wilcoxon Rank Sum **85.8% Power**

Sketchbook

BLINDED WITH SCIENCE

The scientific process, fluid and unpredictable, is one many Americans are never taught. But the pandemic offers a crash course, forcing us to make daily decisions with profound implications amid a flood of often-shifting recommendations. The uncertainty of science can be as infuriating as it is inherent.