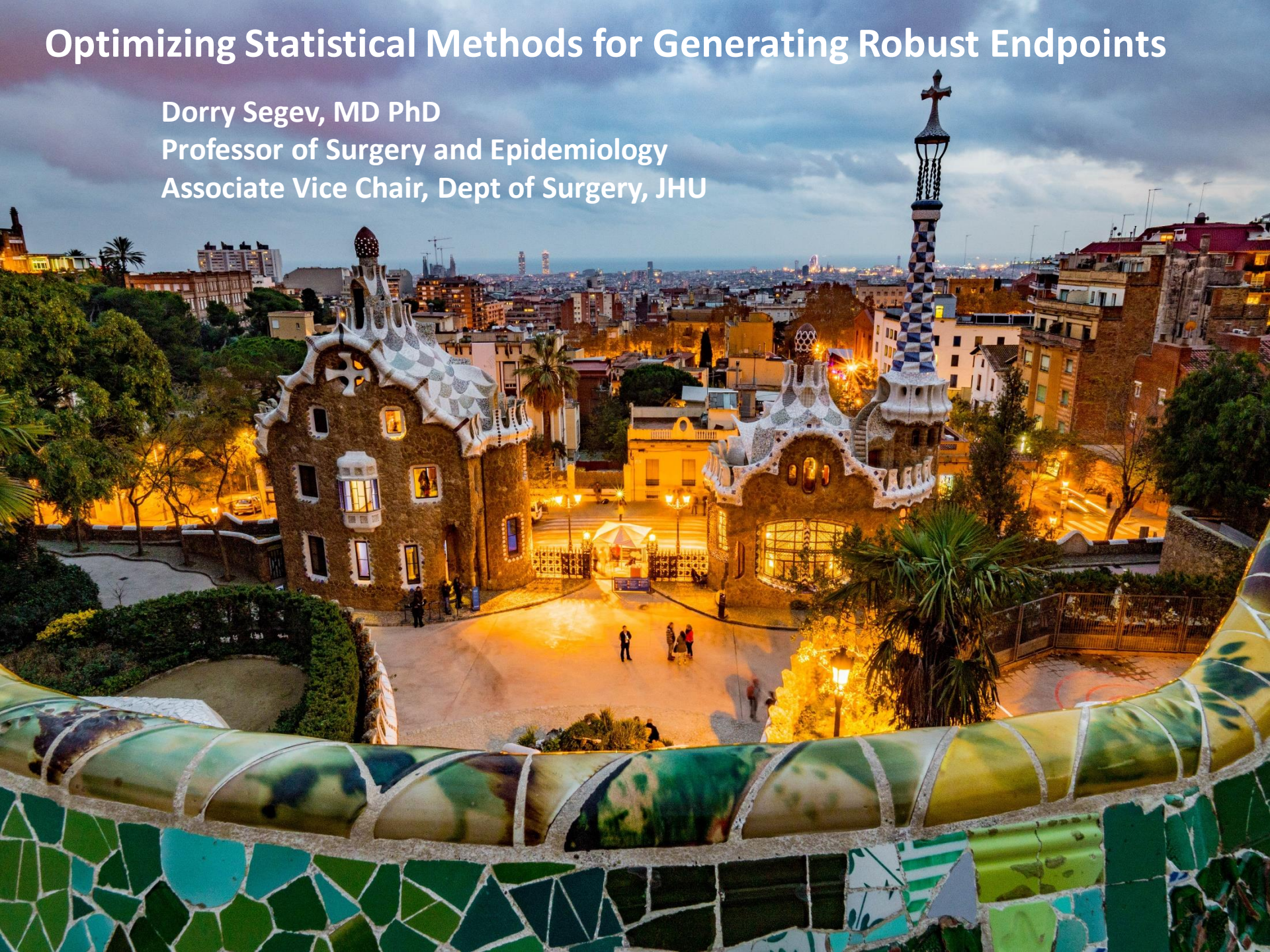
Optimizing Statistical Methods for Generating Robust Endpoints

Dorry Segev, MD PhD
Professor of Surgery and Epidemiology
Associate Vice Chair, Dept of Surgery, JHU



Overview

- Challenges in RCTs
- Standard endpoints: statistical concerns
- Fancier endpoints and other concepts
  - Surrogate endpoints
  - Composite endpoints
  - Adaptive strategies
  - Enrichment strategies

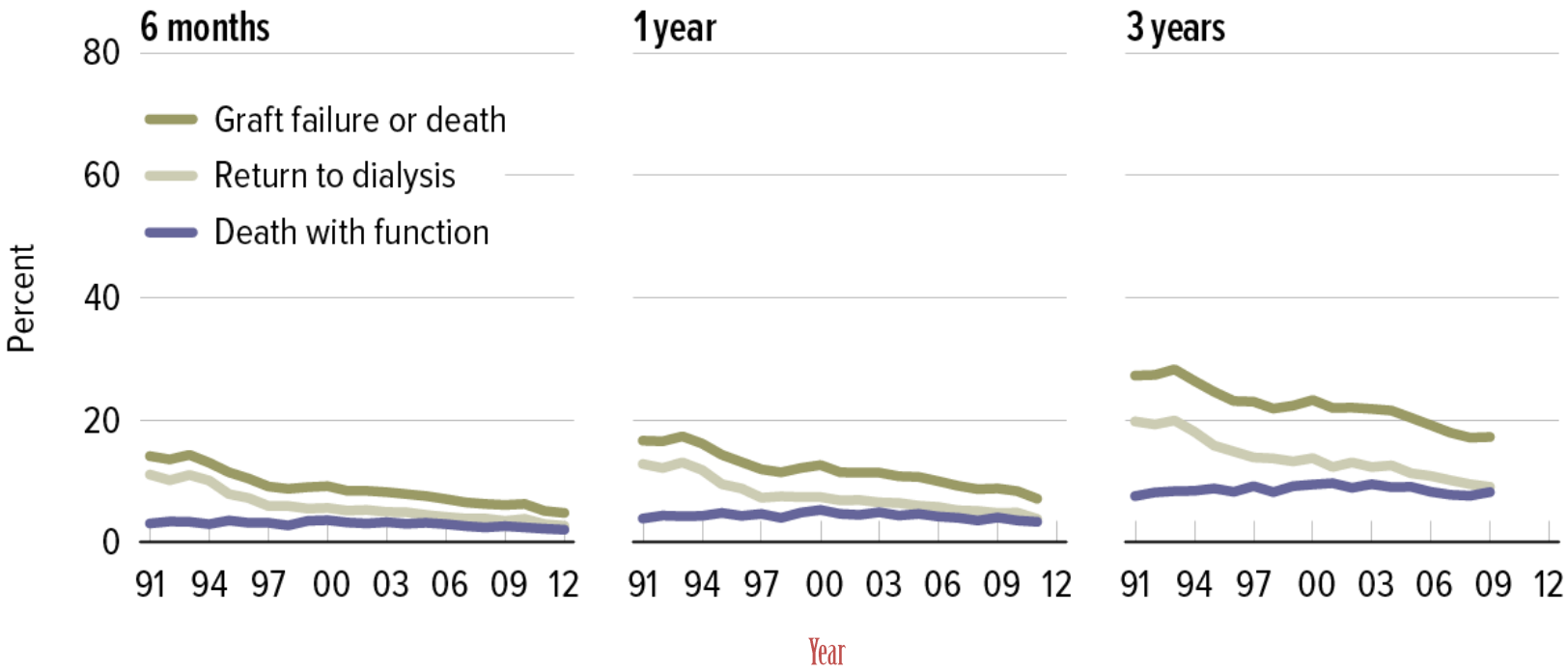
Challenges in RCTs

- Need for large sample size
- Need for long study duration
- Lack of power to evaluate subgroups
- Cost

Goals for Endpoints

- Clinically relevant
- Highly quantitative / easily diagnosible
- Consistently ascertainable in an unbiased way
- Optimally sensitive to treatment
- Precise
- Early response
- Common

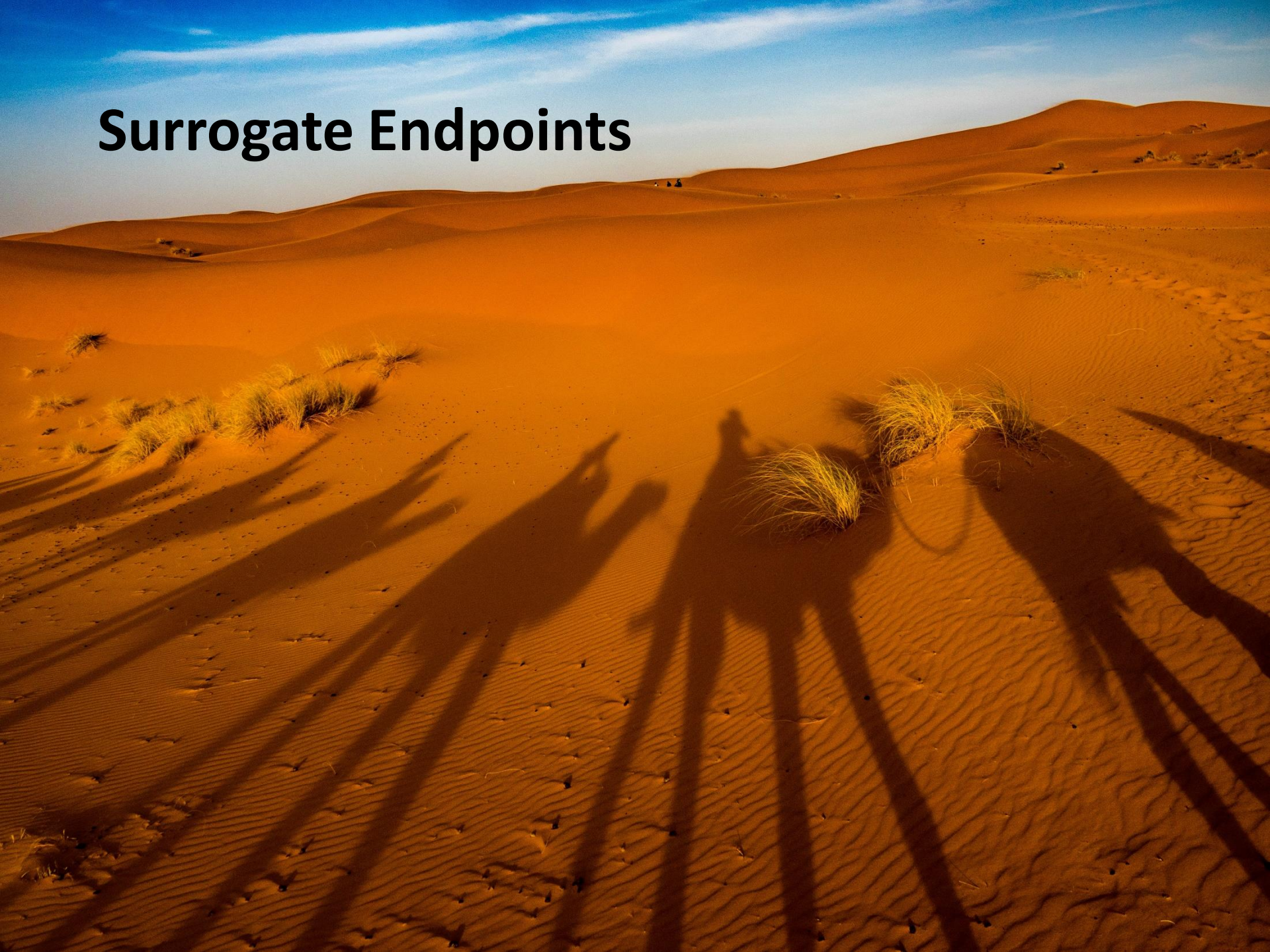
Endpoints in KT: Rare(r) Events



Endpoints in KT: Examples

- Long-term graft loss: long duration, infrequent events
- Short-term graft loss: even more infrequent events
- GFR: measurement error; requires steady state; measured versus estimated; AUC only 0.6 for graft loss; measure absolute or longitudinal? intersubject (between patients) versus intrasubject (within patient)

Surrogate Endpoints



Why Surrogates?

American Journal of Transplantation 2011; 11: 896–906
Wiley Periodicals Inc.

doi: 10.1111/j.1600-6143.2011.03525.x

Meeting Report

Summary of FDA Antibody-Mediated Rejection Workshop

P. Archdeacon[a], M. Chan[b], C. Neuland[c], E. Velidedeoglu[a], J. Meyer[a], L. Tracy[d], M. Cavaille-Coll[a], S. Bala[a], A. Hernandez[a] and R. Albrecht[a]

FDA Definitions

A clinical endpoint (CEP) is an outcome or variable representing a measure of how a patient feels, functions or survives.

In renal transplantation, the current gold-standard CEP is patient and graft survival measured at an appropriate time point.

FDA Definitions

A biomarker is an objectively measured characteristic that is evaluated as an indicator of normal biological or pathogenic processes or pharmacologic responses to an intervention. A biomarker may allow for faster, more efficient clinical trials but greatly depends on the quality of data supporting its use and the setting in which applied.

FDA Definitions

A surrogate endpoint (SEP) is a biomarker intended to substitute for a CEP and expected to predict clinical benefit.

Prentice Criteria

1) The treatment of intervention must affect the surrogate endpoint(s)
2) The treatment or intervention must affect the true endpoint
3) The association of the surrogate endpoint and the true endpoint must be consistent between the treatment or intervention
4) There is an association between the surrogate and the true endpoints

FDA Approval

Prospective RCTs represent a means to evaluate regimens intended to treat AMR. An RCT designed to demonstrate superiority of a treatment on a CEP would demand a large sample size and lengthy follow-up.

While the scientific value of such a study would be significant, it would also be expensive and would take years to complete.

FDA Approval

Given these challenges and the current treatment needs, the [FDA has] discussed the possibility of using an accelerated approval pathway.

FDA Accelerated Approval

A SEP that is 'reasonably likely' to predict a clinical benefit can be used for initial approval. Approval under this regulation requires that the product be studied further after approval to confirm clinical benefit. Extending the pivotal trials which relied on a SEP into the postmarketing period to confirm that the intervention resulted in improved patient and/or graft life, for instance, could represent one possible approach.

Composite Endpoints

Multiple Outcomes

- Over-estimate alpha
- Compensation approaches
  - Hierarchy of outcomes (single primary, multiple secondary)
  - Require stronger statistical evidence
    (eg: $p<0.01$ for 5 outcomes)
  - Convert multiple outcomes to a single one
  - Composite outcome

Why use a composite endpoint?

- **Statistical efficiency**: fewer patients are needed to show a given effect size (there may also be economic/ethical considerations)

Why use a composite endpoint?

- Statistical efficiency
- **Underlying biological considerations**: e.g. graft failure is often a good surrogate for mortality because we may assume that an improvement in graft life will lead to an improvement in patient survival

Why use a composite endpoint?

- Statistical efficiency
- Underlying biological considerations
- **Completeness of drug evaluation:** allows investigators to balance risk and benefit (e.g. include graft rejection, graft loss, and adverse events in a drug trial)

Why use a composite endpoint?

- Statistical efficiency
- Underlying biological considerations
- Completeness of drug evaluation
- **Information preservation**: reduce bias due to informative censoring (e.g. include mortality along with nonfatal events such as rejection, since patients who die may have been at greater risk for rejection)

Types of composite endpoints

- **Binary/time-to-event** – multiple "component" events are combined. If a study subject experiences any of the components, they are considered to have had a failure event
  - Example: [mortality or graft loss]; [AMR or CMR]
- **Continuous** – multiple continuous measures are combined into a single score
  - Example: MELD; Hamilton Depression Rating Scale

Analysis of a composite endpoint

- Only the composite endpoint needs to be formally tested for statistical significance
- All components should trend in the favored direction (e.g. in a trial with acute rejection+mortality as an endpoint, a trend towards greater mortality in the intervention group is worrisome even if not SS)
- This is particularly true if some endpoints (e.g. DGF) are much "softer" than others (e.g. mortality) – composite outcome should not be driven only by the "soft" events

Clinical utility weightings

- If some components are "softer" than others, can pre-specify clinical utility weights to each component (e.g. weight deaths more heavily than DGF episodes)
- However, these weights will be partly arbitrary, and may be disputed

Multiple comparisons

- A single statistical test involving a composite endpoint does not need to be corrected for multiple comparisons
- However, if each endpoint (or a subset of endpoints) will also be tested separately, correction for multiple comparisons is required
- Such secondary endpoints should be prespecified in clinical trials
- Subgroup analyses should also be prespecified

Good Composite Endpoints

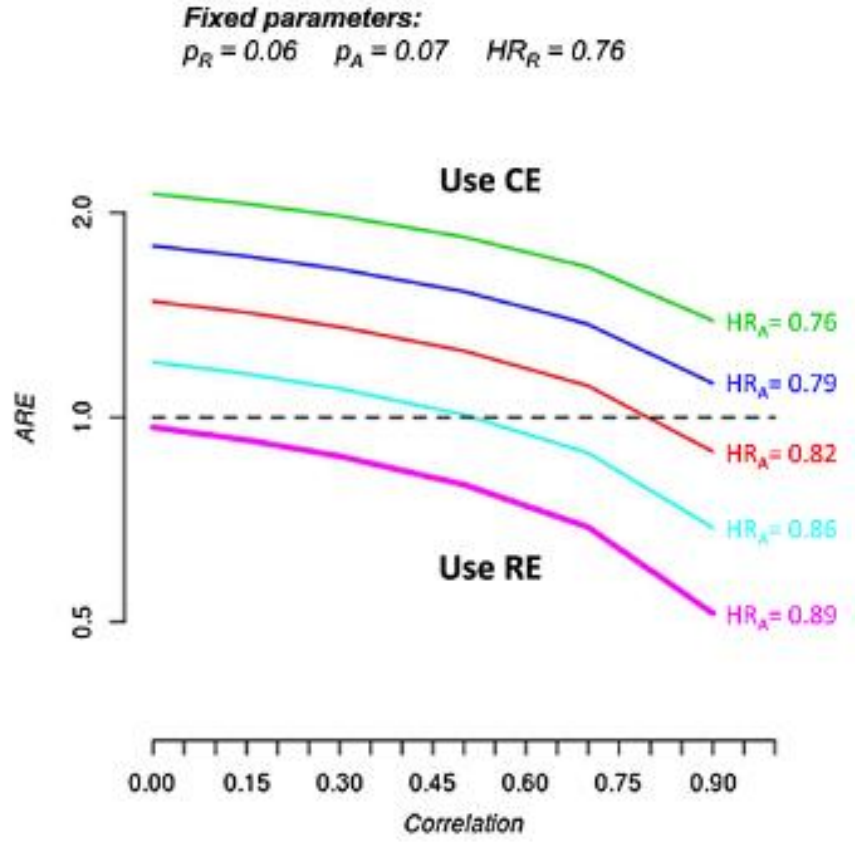
- Individual components are all clinically meaningful to the patient
- Hypothesized associations between the intervention and each component are similar
- Correlation between the components is not too high (higher correlation->less efficiency)

Composite Endpoint = More Power?

- Suppose we have relevant endpoint R and additional endpoint A.
- If the association between exposure and R (HR_R) is greater than the association between exposure and A (HR_A),
- And R and A are strongly correlated…

Composite endpoint = more power? Not always!

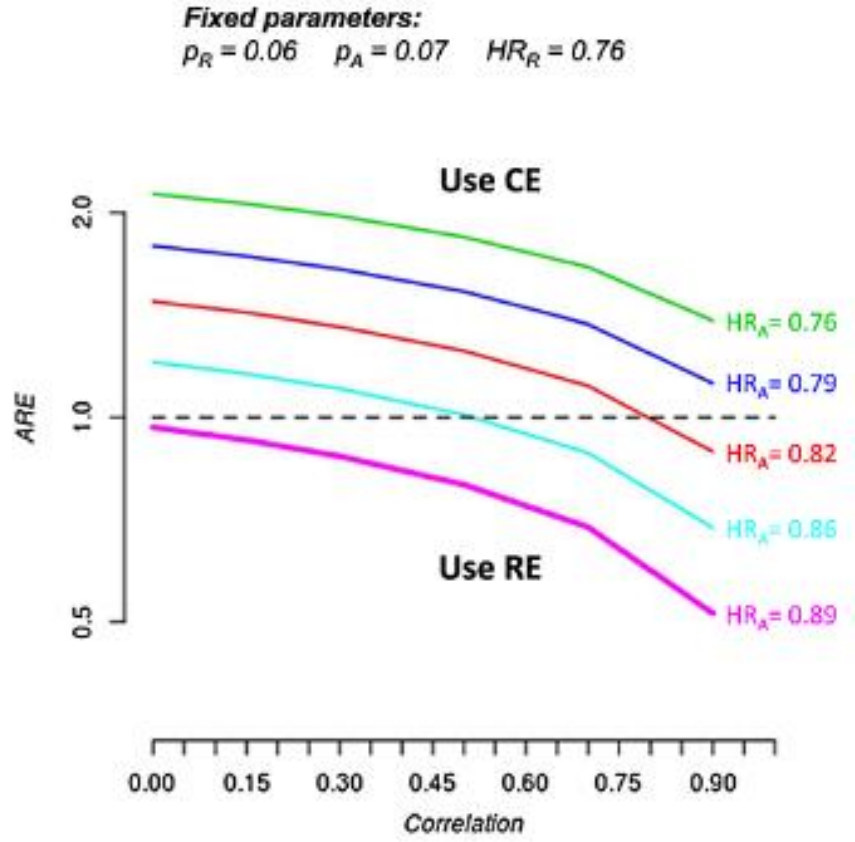
Adding endpoint A may <u>decrease</u> statistical efficiency



Gómez 2015 Stat Med

Composite endpoint = more power? Not always!

If R and A were less strongly correlated, adding A might still increase efficiency



Gómez 2015 Stat Med

Example: HOPE in Action
Endpoint: all-cause graft loss

- Under reasonable assumptions (4.5 events/100 py in control, true HR of HIV+ DD = 1.3) requires **N=200** per group (85% power)
- ACGL does not address main perceived risk of HIV+ DD (HIV progression)

Example: HOPE in Action Endpoint: HIV-Composite

- Composite endpoint of: [acute rejection, graft loss, mortality, AIDS-defining illness, HIV breakthrough]
- Acute rejection and HIV-related outcomes likely have low correlation, improving power
- Acute rejection has higher event rate than other transplant-related outcomes (graft loss, mortality)
- Composite endpoint requires **N=80**

Adaptive Strategies

Challenges in RCTs

- Need for large sample size
- Need for long study duration
- Lack of power to evaluate subgroups
- Cost

Goals of Adaptive Strategies

- Increase trial efficiency
- Potentially benefit trial participants
- Reduce cost
- Enhance likelihood of finding a true benefit (if exists)

Adaptive Strategies: Exploratory

- Exploratory trials:
  - Finding safe and effective doses
  - Dose-response modeling
- Adaptive strategy goals:
  - Assign a larger proportion of participants to treatment groups that are performing well
  - Reduce number of participants in treatment groups that are performing poorly
  - Investigate a larger dose range (than nonadaptive designs)

Adaptive Strategies: Confirmatory

- Adaptive strategy goals:
  - Make prospectively planned changes to the future course of an ongoing trial on the basis of an analysis of accumulating data from the trial itself (blinded or unblinded) *without undermining the statistical validity of the conclusions*
  - Need to make sure implementation is scientific, ethical, free from bias (especially unblinded)

Adaptive: 2-Stage Design

- Stage 1 is broad dose testing and dose selection by DSMB for stage 2
- *Final analysis can use observations from both stages (but is complex and uses non-conventional parameter and confidence interval estimates)*
- Requires early (short-term) endpoint
- No sponsor involvement in dose selection
- Risk of inadequate dose–response modeling

Adaptive: Sample-Size Reestimation

- At x% (70%) enrollment, effect size estimated and binned: unfavorable, promising, favorable zones
- N increased if in promising zone (in traditional, fix N to maximum amount of information)
- Small loss of overall power (interim analysis), substantial gain in conditional power (if promising)
- Particularly useful when effect size unknown
- Requires meticulous planning, operationalization

Adaptive: Changing Endpoint

- Start as noninferiority trial
- Based on interim analysis, continue to enroll to show superiority if conditional power to show superiority exceeds a given cutoff
- Can use patients from noninferiority study in final superiority analysis

Adaptive Strategies: Requirements

- Meticulous preliminary planning
- Detailed adaptive criteria
- Dissemination of interim results without unblinding interim results
- Hypothesis testing strategy to control type I error (small loss of power at each interim)
- Detailed simulations before initiation

Enrichment Strategies

Subject Selection

Tight criteria

- Reduces variability and sample size
- Excludes subjects at risk of treatment complications
- Includes subjects most likely to benefit
- May restrict to advance disease, compliant patients, etc.
- Slows enrollment
- “Best case” results

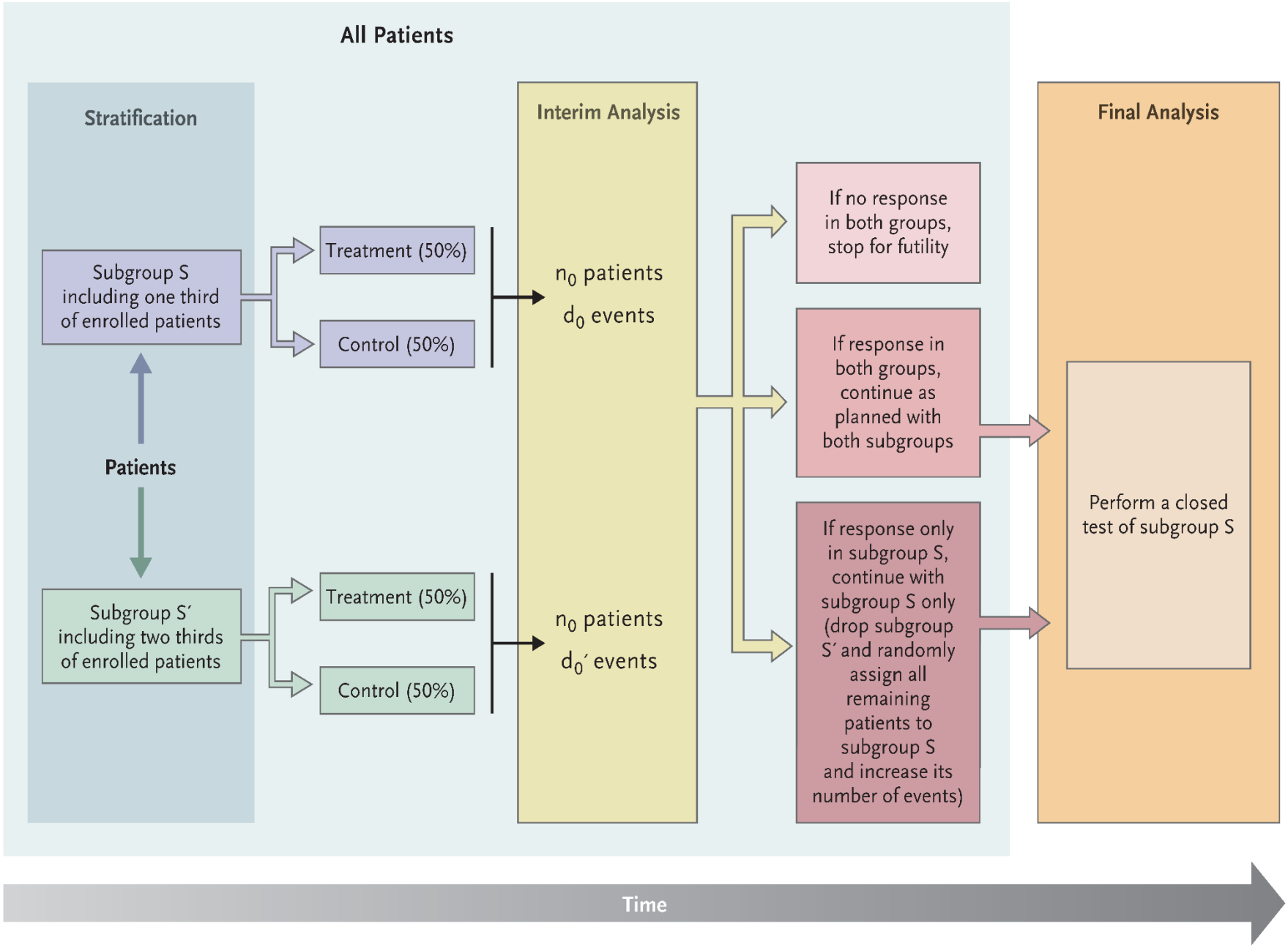
Subject Selection

Loose criteria

- Increases variability and sample size
- Speeds enrollment
- Enhances generalizability
- “Real world” participants

Enrichment Strategies

- (Biomarker-driven) population-enrichment
- Considers treatment effect heterogeneity
- Enrichment Adaptive = (1) study whether a given profile is predictive for success of therapy and (2) enrich population to those most likely to benefit
  - Initially randomize regardless of profile, then interim analysis tests effect modification, and then possibly terminate enrollment for some profiles
  - Final analysis incorporates data from both phases


All Patients
Stratification
Patients
Subgroup S including one third of enrolled patients
Subgroup S′ including two thirds of enrolled patients
Treatment (50%)
Control (50%)
Treatment (50%)
Control (50%)
Interim Analysis
n_0 patients
d_0 events
n_0 patients
d_0' events
If no response in both groups, stop for futility
If response in both groups, continue as planned with both subgroups
If response only in subgroup S, continue with subgroup S only (drop subgroup S′ and randomly assign all remaining patients to subgroup S and increase its number of events)
Final Analysis
Perform a closed test of subgroup S
Time

Summary

- Challenges in RCTs
- Standard endpoints: statistical concerns
- Fancier endpoints and other concepts
  - Surrogate endpoints
  - Composite endpoints
  - Adaptive strategies
  - Enrichment strategies