Utilities and Pitfalls of Composite and Surrogate Endpoints in Clinical Trials

Helen Parise, ScD

Cardiovascular Research Foundation
Columbia University Medical Center

CARDIOVASCULAR RESEARCH FOUNDATION

COLUMBIA UNIVERSITY MEDICAL CENTER

Common CV Clinical Endpoints

- Death (Cardiac, Non-Cardiac)
- Myocardial Infarction (Q-Wave, NQWMI)
- Stent Thrombosis (subcategories)
- Target Lesion Revascularization
- Target Vessel Revascularization

Surrogate Endpoints

- **What is a surrogate endpoint?**
  - An outcome measure which may occur sooner or with greater frequency than the "true" outcome of interest
  - Predictive of the "true" clinical endpoint
  - In the causal pathway

Treatment → Surrogate → Clinical Endpoint

Surrogate Endpoints in DES Studies

- Surrogates in DES Studies
  - Usually continuous variables that are objectively quantifiable and reproducible
- IVUS: Neointimal Volume
- Angiographic:
  - Late Lumen Loss
  - Percent Diameter Stenosis

Surrogate Endpoints

- Why use surrogate endpoints?
  - Reduce sample size with adequate power
    - fewer patients + shorter follow-up = reduced costs
  - Avoid/minimize randomizing patients to a therapy which might not provide benefit or possibly cause harm
  - Test new technologies and be able to anticipate their outcomes

Surrogate Endpoints

- How do I know if a surrogate is good enough to use in my clinical trial?

Statistical Criteria for Evaluating a Surrogate

- Access to patient level data from multiple well conducted trials is required
- Analysis of 11 trials evaluating DES v. BMS
  - TAXUS-IV, TAXUS-V, TAXUS-VI, SIRIUS, E-SIRIUS, C-SIRIUS, RAVEL, DELIVER, REALITY, ENDEAVOR II, ENDEAVOR III
- In-Segment and In-Stent Late Loss (LL) and %DS were evaluated

Statistical Criteria for Evaluating a Surrogate

1. Consistent evidence of treatment differences within each trial

Pocock, et al., JACC 2008; 51 p 23-32

1. Evidence of treatment difference

- *Z*-Scores for Treatment Difference

	TAX IV	TAX V	TAX VI	SIR-IUS	E-SIR-IUS	C-SIR-IUS	RA-VEL	DE-LIV-ER	RE-AL-ITY	END II	END III
In-Stent											
LL	11.6	10.5	10.6	18.7	14.5	8.2	13.7	3	7.4	9.2	8.3
%DS	12.5	10	10.5	18.5	14	8.1	11.9	2.7	3.3	8.5	6.5
In-Segment											
LL	8.9	7.3	7.2	13.1	11.1	5.8	9.8	2.3	4.5	7.7	4.4
%DS	9.4	6.2	8.2	13.3	11	6.8	7.2	2.3	2.5	7.3	3.4
TLR	4.7	4	3.6	7.1	6.1	3.1	3.8	1.3	1.1	4	1.1

Statistical Criteria for Evaluating a Surrogate

1. Consistent evidence of treatment differences within each trial
2. Strong relationship with clinical outcome

Pocock, et al., JACC 2008; 51 p 23-32

2. Relationship to clinical outcome

- c-statistics summarizing the strength of association to TLR in each trial

	In-Stent		In-Segment			In-Stent		In-Segment	
Trial	LL	%DS	LL	%DS	Trial	LL	%DS	LL	%DS
TAXUS-IV	0.9	0.94	0.92	0.97	RAVEL	0.98	0.93	0.91	0.95
TAXUS-V	0.88	0.9	0.91	0.95	DELIVER	0.85	0.86	0.86	0.91
TAXUS-VI	0.86	0.87	0.89	0.95	REALITY	0.9	0.93	0.93	0.98
SIRIUS	0.88	0.9	0.92	0.95	END. II	0.88	0.91	0.88	0.95
E-SIRIUS	0.94	0.95	0.92	0.95	END. III	0.77	0.8	0.93	0.95
C-SIRIUS	0.86	0.86	0.93	0.94	Averaged	0.88	0.9	0.91	0.95

CARDIOVASCULAR RESEARCH FOUNDATION

Pocock, et al., JACC 2008; 51 p 23-32

COLUMBIA UNIVERSITY MEDICAL CENTER

Statistical Criteria for Evaluating a Surrogate

1. Consistent evidence of treatment differences within each trial
2. Strong relationship with clinical outcome
3. Treatment difference in clinical outcome statistically explained by the surrogate *within* each trial [Prentice criterion]

Pocock, et al., JACC 2008; 51 p 23-32

3. Treatment Differences explained within trial [Prentice criterion]

- Percent treatment effect explained* (largest 5 trials shown)

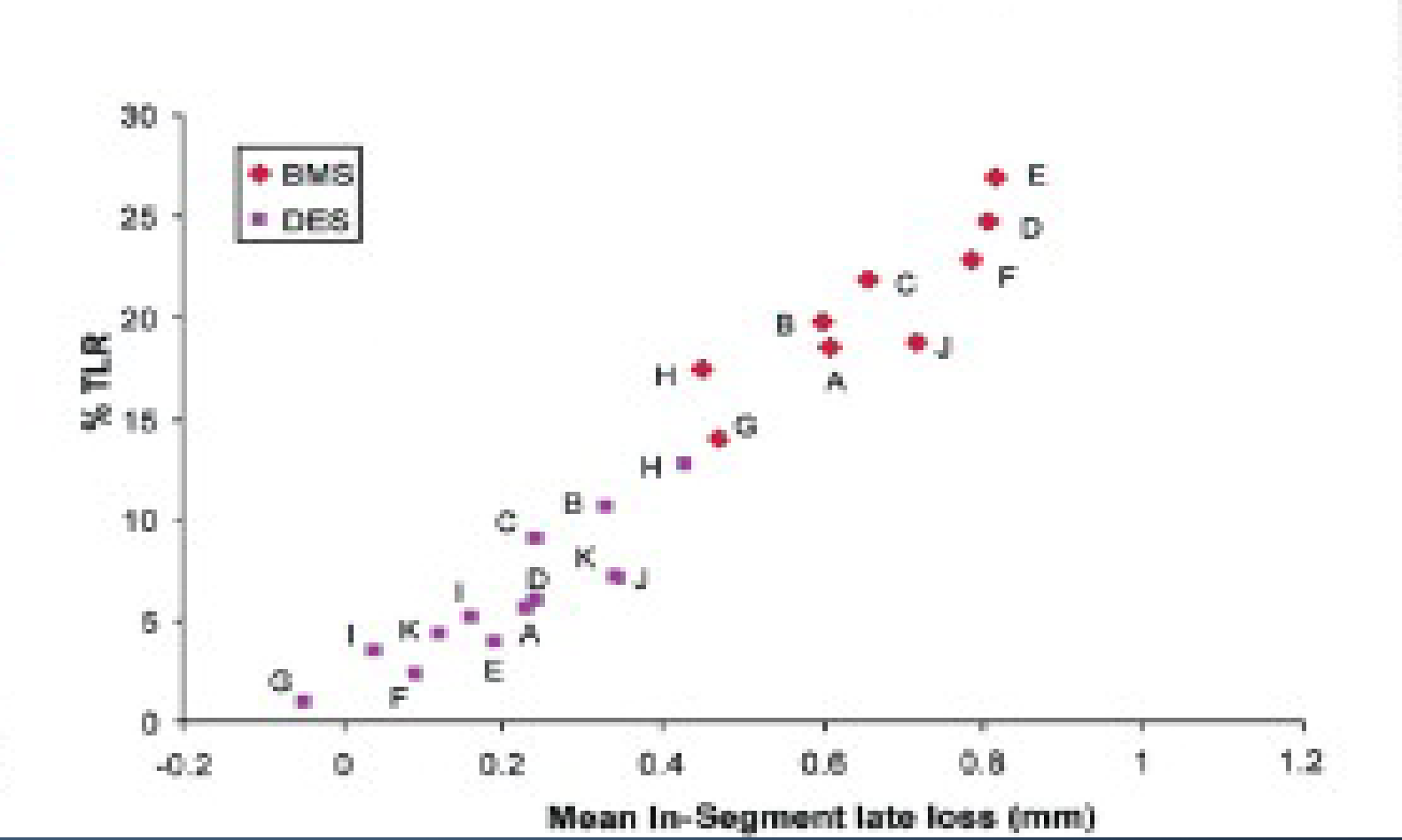
	In-Stent		In-Segment	
Trial	**LL**	**%DS**	**LL**	**%DS**
TAXUS-IV	98	138	83	83
TAXUS-V	101	94	82	47
TAXUS-VI	115	124	75	114
SIRIUS	132	142	91	103
ENDEAVOR II	100	119	82	107

* from logistic model for log odds of TLR after potential surrogate (e.g., LL) is added as predictor variable.

Statistical Criteria for Evaluating a Surrogate

1. Consistent evidence of treatment differences within each trial
2. Strong relationship with clinical outcome
3. Treatment difference in clinical outcome statistically explained by the surrogate *within* each trial [Prentice criterion]
4. Magnitude of treatment difference in clinical outcome clearly linked to magnitude of treatment difference in surrogate *across* trials [Hughes criterion]

4. Relating size of treatment effect across trials [Hughes criterion]



Pocock, et al., JACC 2008; 51 p 23-32

Example: XIENCE V

- SPIRIT III trial
- Primary endpoint: in-segment late loss at 8 months

	Xience V	Taxus	
N	669	332	
Mean Late Loss	0.14mm	0.26mm	P=0.003
Target Lesion Revasc.	3.4%	5.6%	P=0.12
1 year MACE	6.0%	10.3%	P=0.02

- Larger long term trials (and registries) underway to assess patient safety re death, MI, stent thrombosis

Stone, et al. JAMA 2008; 299 p 1903-1913

Surrogates

- Reduced trial size → more efficient progress
- LL and % DS both look good as surrogates for TLR in DES
- Use can speed up the regulatory process …

  IF the surrogate is acceptable to the regulatory agency

Surrogate Endpoints – Some Pitfalls

- May not predict clinical events in “real world”
- Surrogates for efficacy may not (and likely do not) predict safety
- Can lead to premature approval of potentially unsafe treatment
- Missing data issues: e.g. incomplete follow-up in angiographic endpoints may bias results

Composite Endpoints

- **What is a Composite Endpoint?**
  - Common in Cardiovascular Trials
  - Usually a group of clinical outcomes considered together to form a single endpoint
  - Generally weighted equally
  - Experiencing any of the individual outcomes → experiencing the composite

Composite Endpoints: Examples

- Examples of Composite Endpoints used in Cardiovascular Trials
  - Cardiac death/MI
  - Device-oriented (TLF): cardiac death/MI/TLR
  - TVF: cardiac death/MI/TVR
  - MACE: a mixed bag (and varies from study to study!)

Composite Endpoints

- Why use Composite Endpoints?
  - Gain in statistical power
  - Simple summary of several outcomes
  - Captures broader range of treatment experience
  - Multiple outcomes tested simultaneously without alpha “hit” for multiplicity
  - Avoids (some) issues of competing risks

Composite Endpoints

- Considerations
  - All components should be affected equally by treatment otherwise composite may be diluted
  - All components should be clinically relevant
  - All components should be defined as secondary and reported separately

Composite Endpoints

- In a review of 14 journals between 1 Jan 2000 and 1 Jan 2007
  - Of all cardiovascular trials (1231), 37% used composite endpoints
  - 98% of these included Mortality as a component
  - Typically 3 to 4 outcomes were included (based on the IQR)

Lim, et al. Ann Intern Med. 2008;149:612-617

Example: SYNTAX

"SYNTAX trial fails to show non-inferiority for DES" - Cardiovascular News, Sept 2008

- 1800 patients with left main/3 vessel disease comparing CABG vs. TAXUS
- MACCE = death/stroke/MI/Revascularizations
- 1-Year results were complex
- QOL results presented at ACC '09

Example: SYNTAX (1-year Results)

	CABG	TAXUS	
N Randomized	897	903	
MACCE (primary)	105 (12.4%)	159 (17.8%)	P=.002
Death	30 (3.5%)	39 (4.4%)	
Stroke	19 (2.2%)	5 (0.6%)	P=.003
MI	28 (3.3%)	43 (4.8%)	
Death/MI/Stroke	65 (7.7%)	68 (7.6%)	P=.98
Revascularization	50 (5.9%)	120 (13.5%)	P<.0001
PCI	40 (4.7%)	102 (11.4%)	
CABG	11 (1.3%)	25 (2.8%)	

Serruys, et al. N Engl J Med 2009; 360:961-972

Example: HORIZONS (30 day Results)

- **Evaluating efficacy and safety in composite endpoint**

	Heparin +IIb/IIIa	Bivalirudin alone	
N	1802	1800	
Major Bleed	8.3%	4.9%	P<0.001
MACE*	5.5%	5.4%	
Net Adverse Clinical Events	12.1%	9.2%	P=0.005
Death	3.1%	2.1%	P=0.047

* MACE = Death/MI/Stroke/TLR

Composite Endpoints: Some Pitfalls

- Survival analysis may count less severe outcomes over more severe (e.g. TLR vs. Mortality)

  Example: MACE = Death/MI/Stroke/TLR

  Patient A has TLR at 5 months but survives and feels well through end of trial (5 years)

  → Time to MACE = 5 months

  Patient B has no TLR, MI or Stroke, but dies at 9 months → Time to MACE = 9 months

Q: Is patient A really worse off than patient B?

Composite Endpoints: Some Pitfalls

- Can be misleading
  - MACE = Death/MI/Stroke/TLR
  - Why not

    MACE = TLR/MI/Death/Stroke?
  - Most important component listed first, but often has the lowest event rate (i.e. is the LEAST represented in the composite)

Composite Endpoints: Some Pitfalls

- All components tend to be treated equally
  - Can be weighted (but what are correct weights?)
- Some might presume that benefits relate to ALL components
- May dilute real treatment differences by including elements unaffected by treatment (impact on non-inferiority?)
- Can be difficult to interpret

CONCLUSIONS

- Both surrogates and composite endpoints are extremely useful for reducing sample sizes (increasing power)
- Both can get “better” treatments to patients faster
- But …
- Proceed with CAUTION!