The Consumer Reports Guide to Trials

Trials on Trial: What Can Be Done to Preserve and Sustain RCTS in the Future

AHA/QCOR Plenary Session, 2016

Phoenix, AZ

Sanjay Kaul, MD, FACC, FAHA

Division of Cardiology

Cedars-Sinai Medical Center

Los Angeles, California

CEDARS-SINAI

Duality of Interests

- Advisor/Consultant
  - United States FDA
- Consultant
  - Boehringer Ingelheim
  - AstraZeneca
  - Takeda
- Stocks/Equity
  - Johnson & Johnson

The Laws of Diminishing Objectivity in the Interpretation of Evidence

- vehemence α evidence^{-1}

- vehemence α eminence^{2}

Peter McCulloch
The Lancet, 2004;363;9004

Randomized Controlled Trial (RCT)

- Greatest medical invention ever
  - Methodological paragon for assessing evidence
- Randomization ensures similar groups at start and assignment is fair
  - Balances measured and unmeasured covariates
- Double blinding ensures a level playing field
  - Can’t favor one arm over the other
- Incentives encourage rigorous study conduct
  - Sloppiness makes arms more similar

Hallmarks of a Good Randomized Controlled Trial

- Random concealed allocation
- Double blinding
- Intention to treat analysis
- Simple, large scale and pragmatic
- Unrestricted patient population
- Adequate power to minimize false positive & negative errors
- Easily ascertained, clinically important, hard endpoints
- Risk benefit analysis
- Cost effectiveness analysis

Evidence Standards

FDA/CDER	“Substantial evidence of Effectiveness”
FDA/CDRH (PMA)	“Reasonable assurance of Safety and Efficacy”
FDA/CDRH (510k)	“Substantial evidence of Equivalence”
ACC/AHA	“Useful and Effective”
CMS	“Reasonable and Necessary”
Payers	“Usual and Customary”
Courts	“Prudent and Cautious”
Consumer Reports	*“Reliable and meaningful”*

Standards for Developing Trustworthy Clinical Practice Guidelines: IOM Report

- Establishing Evidence Foundations (Standard 5)
  - *Quality of evidence*
  - *Quantity of evidence (magnitude and precision)*
  - Consistency of aggregate available evidence
  - *Clear description of benefits and harms*
  - Rating of strength of recommendation

http://www.iom.edu/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx

Quantity of Evidence Necessary to Support Effectiveness of Drugs and Biologics: FDA

CFR	Statutory criterion
FDC Act 1962	"Substantial" evidence of effectiveness consisting of "adequate and well-controlled investigation**s**", i.e., two separate trials each with $p<0.05$ ($0.05 \times 0.05 = 0.0025$ divided by 2 = 0.001)
FDA Evidence Guidance for Industry, 1998	"A highly persuasive statistical finding (a p value <0.001) in a single trial with some other indication of the study's reliability (e.g., multicenter with no center driving the results)"
FDAMA 115 (1998)	"One adequate and well-controlled study and confirmatory evidence."

Replication Probability and P-values

P value of initial expt.	Probability of p<0.05 when the first observed difference is true
0.10	.37
0.05	.50
0.03	.58
0.01	.73
0.005	.80
0.001	.91

‘Replication is at the heart of scientific endeavor’

Goodman, SN, “A Comment on Replication, P-values and Evidence, *Stat Med*, 11:875-879, 1992.

Foundations of Classical Statistical Inference

PURSUIT Trial (Death or MI)

Study Group	OUTCOME +	EVENT -	Total
Ept	671	4051	4722
Con	744	3995	4739
Total	1415	8046	9461

	N	Event Rate (%)
Eptifibatide	4722	14.2
Control	4739	15.7

RRR	95% CI	P (2-tailed)
9%	1%-18%	0.04

Objective:

- Evaluate impact of eptifibatide on adverse cardiac outcomes in patients with NSTE ACS

Sample size estimation:

- Study powered to detect a minimum clinically important difference (δ) of 20% risk reduction

Conclusion:

- Eptifibatide is superior to aspirin and heparin in NSTE ACS

Does the evidence justify strong recommendation?

What Does a P(ee) Value of 0.05 Mean?

- 'Fisherian' P value of 0.05 is arbitrary and originally based on n=30!
- Always demand a P value of <0.001 for a sample size > 200 as strong evidence against the null hypothesis of zero difference

Al Feinstein

Bayes Factor
Quantifying Strength of Evidence

- Bayes' theorem (Reverend Thomas Bayes, 1763)

  Posterior odds = prior odds x evidence (Bayes factor)

- Bayes factor
  - BF = Prob (Data/H_0)/Prob (Data/H_1) (likelihood ratio)
  - H_0 = Null hypothesis; H_1 = alternative hypothesis
  - Odds = Probability/(1-Probability)
  - Probability = Odds/(1+Odds)
  - Minimum BF =exp(-0.5z^2)

Bayes factor is a comparison of how well two hypotheses predict the data: smaller the BF, stronger the evidence against H_0

Evaluating Strength of Evidence by Bayes Factor

PURSUIT (Death or MI)

- Minimum Bayes factor
  - z = 2.032
  - P value = 0.04
  - Minimum BF =exp(-0.5z^2) = 0.13
  - Prior null probability = 0.50
  - Prior null odds = 0.50/(1-0.50) = 1
  - Posterior null odds = 1 x 0.13 = 0.13
  - Posterior null probability = 0.13/(1+0.13) = 0.11

P Value (Z score)	Minimum Bayes Factor	Decrease in Probability of Null Hypothesis, %		Strength of Evidence
		From	To No Less Than	
0.04 (2.03)	0.13 (1/7.7)	75 50 25	28 11 4	Moderate

Relationship Between P values & Bayes Factor

P Value (Z score)	Minimum Bayes Factor	Decrease in Probability of Null Hypothesis, %		Strength of Evidence
		From	To No Less Than	
0.10 (1.64)	0.26 (1/3.8)	75 50 17	44 21 5	Weak
0.05 (1.96)	0.15 (1/6.8)	75 50 26	31 13 5	Moderate
0.03 (2.17)	0.095 (1/11)	75 50 33	22 9 5	Moderate
0.01 (2.58)	0.036 (1/28)	75 50 60	10 3.5 5	Moderate to Strong
0.001 (3.28)	0.005 (1/216)	75 50 92	1 0.5 5	Strong to very strong

- *P value overstates the strength of evidence against null hypothesis*
- *As the strength of evidence increases (smaller P value), the discrepancy between P value and Bayes factor becomes negligible*

“There is a tendency to make the measurable important, rather than the important measurable”

Robert S. McNamara

Statistical Significance vs. Clinical Importance

- MDD (minimum detectable difference, "δ")
  - The "minimum difference" the study is powered to detect
  - Utilized for sample size estimation
  - May or may not reflect a clinically important difference
    (driven by financial constraints, restricted availability and follow up, etc)

- MCID (minimum clinically important difference)

  The "minimum acceptable difference" to change the behavior of the clinician, patient, payer or policy maker, given the side effects, costs and inconveniences of therapeutic interventions

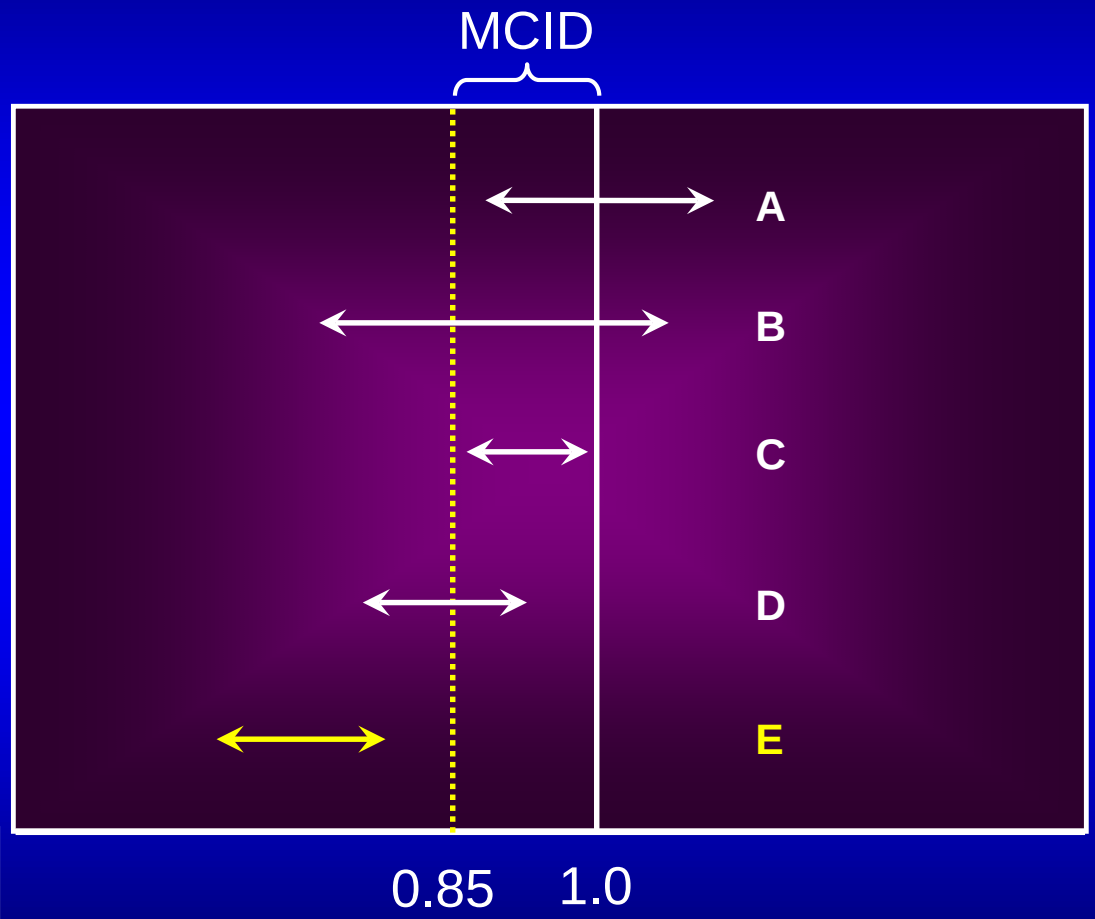
Statistical Significance vs. Clinical Importance
MCID Threshold for UA/NSTEMI ACS

"In ACS, a relative reduction of 15% in recurrent clinical events has recently been considered clinically important (GUSTO I); this level is far below the perceived threshold that drove the sample size calculations for clinical trials just a decade ago. As we develop more incrementally beneficial therapies, it is likely that the minimally important clinical difference will become even smaller."

Califf and DeMets
Circulation. 2002;106:1015

Statistical Significance vs. Clinical Importance

Strength of Evidence



Statistically not significant, clinically not important

Statistically not significant, may be clinically important

Statistically significant, not clinically important

Statistically significant, may be clinically important

Statistically significant, clinically important

MCID = minimal clinically important difference = 15% RRD

Sackett, D

Statistical Significance vs. Clinical Importance

Class I, LOE A Recommendations for UA/NSTEMI

Impact on Death or MI

Intervention	Control (%)	Rx (%)	Summary risk ratio (95% CI)	P Value	NNT (95% CI)	Interpretation of Confidence Intervals (MCID = 15% RRR)
Aspirin (N=2,856)	**12.8**	**5.5**	**0.43 (0.33-0.56)**	**<0.01**	**14 (11-19)**	**Statistically significant and clinically important (E)**
UFH (N=1,353)	10.4	7.9	0.67 (0.44-1.02)	0.06	44 (∞-18)	Statistically not significant, maybe clinically important (B)
Enoxaparin (Early invasive)	12.8	12.1	0.96 (0.88-1.05)	0.35	171 (∞-59)	Statistically not significant, clinically not important (A)
Clopidogrel (CURE)	11.4	9.3	0.82 (0.74-0.92)	<0.01	54 (35-120)	Statistically significant, maybe clinically important (D)
GP IIb/IIIa (Early invasive)	14.5	11.8	0.81 (0.70-0.94)	0.007	37 (21-139)	Statistically significant, maybe clinically important (D)

Aspirin is the only intervention listed as a performance measure!

Judging the Strength of Evidence
Summary

- Statistical significance tells us whether a difference is likely to be real (P value)
- Clinical importance tells us whether the difference is small or large, trivial or important ("oomph" value)
- Guidelines currently emphasize statistical significance over clinical importance
- Ideally, assessment of both statistical significance and clinical importance should aid in optimal utilization of therapeutic interventions in clinical practice

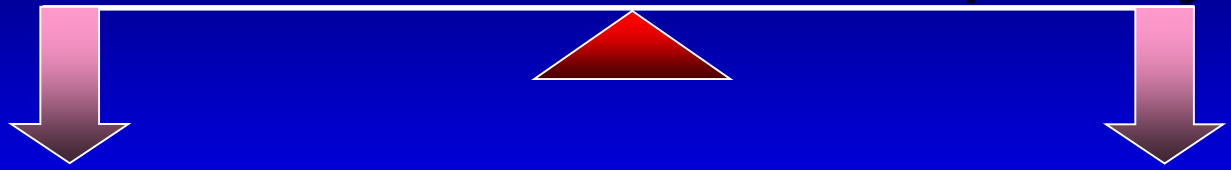
Standards for Developing Trustworthy Clinical Practice Guidelines: IOM Report

- Establishing Evidence Foundations (Standard 5)
  - *Quality of evidence*
  - *Quantity of evidence (magnitude and precision)*
  - Consistency of aggregate available evidence
  - *Clear description of benefits and harms*
  - Rating of strength of recommendation

http://www.iom.edu/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx

Factors that Modify the Quality of Evidence

And Should be Considered Explicitly in CPGs



↓ Quality

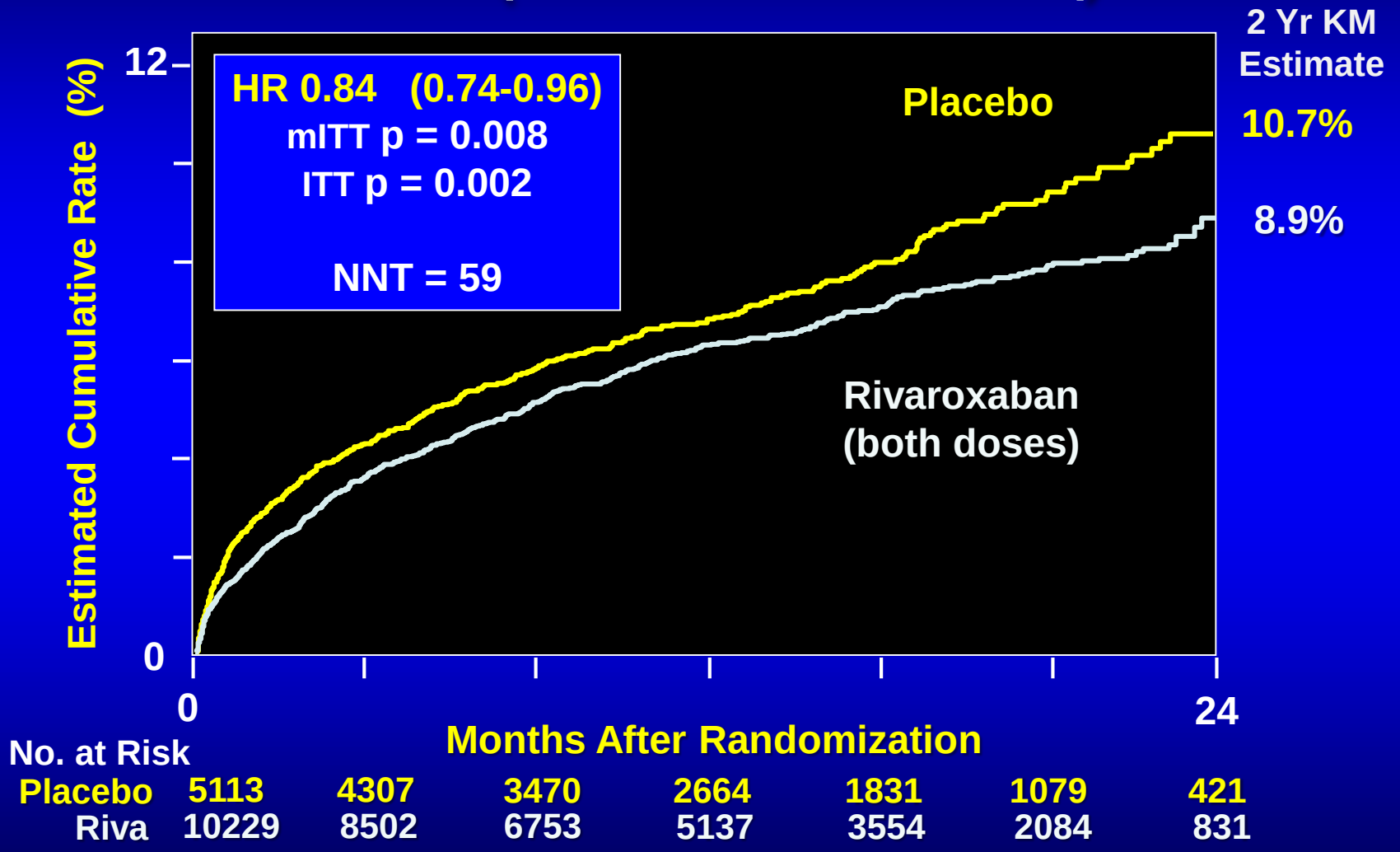
- Important limitations
  - Study design or execution (bias)
    - Randomization
    - Lack of concealment
    - Inadequate blinding
    - ITT principle violated
    - Loss to follow-up
    - Early stopping for benefit
  - Indirectness of results
  - Inconsistency of results
  - Imprecision
  - Publication bias

↑ Quality

- Special strengths
  - Magnitude of effect
    - RR<0.5 or >2.0 (large)
    - RR<0.2 or >5.0 (very large)
  - Dose-response gradient
  - All plausible confounding would underestimate effect or all plausible biases would overestimate effect

ATLAS ACS-2 TIMI 51

Rivaroxaban (Anti-Xa Inhibitor) in ACS



Mega et al, NEJM 2012;366:9-19

Missing Data in Contemporary ACS Trials
ATLAS ACS-2 TIMI 51

Trial Name	Study Agent	Enrolled (N)	Follow-up (Median)	Incomplete follow up N (%)	Withdrawal of consent N (%)	Vital status unknown N (%)
ATLAS	Rivaroxaban	15,526	484 d	**2402 (15.5)**	**1294 (8.9)**	**1117 (7.2)**
APPRAISE-2	Apixiban	7,392	241 d	131 (1.8)	81 (1.1)	NR
TRACER	Vorapaxar	12,944	502 d	761 (5.9)	NR	249 (1.9)
PLATO	Ticagrelor	18,624	277 d	562 (3.0)	545 (2.9)	2 (0.01%)
TRITON	Prasugrel	13,619	14.5 m	804 (5.9)	665 (4.9)	16 (0.12)

- Relatively high rate of withdrawal of consent & missing vital status in ATLAS
- Differential dropout for MACE (12.4% Riva vs. 11% placebo)

Krantz M and Kaul S, JACC 2013

Impact of Missing Data

Sensitivity Analysis to Assess 'Missingness' Tolerability

Group	Stratum	Rivaroxaban	Placebo	Relative Risk (95% CI)	Excess # event , Riva
Pooled	All	626/10229 (6.1) 666/10229 (6.5)	376/5113 (7.4) 376/5113 (7.4)	0.84 (0.74-0.96) 0..89 (0.78-1.00)	40
	2 (+ DAPT)	575/9532 (6.0) 598/9532 (6.3)	340/4760 (7.1) 340/4760 (7.1)	0.86 (0.75-0.98) 0.88 (0.77-1.00)	23
2.5 mg	All	313/5114 (6.1) 326/10229 (6.4)	376/5113 (7.4) 376/5113 (7.4)	0.84 (0.72-0.97) 0.87 (0.75-1.00)	13
	2 (+ DAPT)	286/4765(6.0) 293/4765 (6.2)	340/4760 (7.1) 340/4760 (7.1)	0.85 (0.72-0.99) 0.86 (0.74-1.00)	7

Number of potential excess events in Rivaroxaban arm to overturn significance relatively low, indicating fragility of treatment benefit (intolerable 'missingness')

Krantz M and Kaul S, JACC 2013

Impact of Missing Data
Sensitivity Analysis in ATLAS ACS-2 TIMI 51

Incomplete follow up for Rivaroxaban:	**2192**
Excess Rivaroxaban events to p>0.05:	**7 to 40**
% Patients with TIMI major/minor bleeds who experience MACE:	**37%**
No. of TIMI major minor bleeds to yield 7-40 excess MACE (7/0.37 to 40/0.37)	**19 to 108**
No. of Rivaroxaban incomplete f/u, no MACE, but TIMI major/minor bleed	**98**

More bleeding with Rivaroxaban & incomplete f/u can explain MACE difference

FDA Briefing Document 2012

Missing Data & Informative Censoring
TIMI Major or Minor Bleeding in ATLAS

Group	Incomplete f/u (per 100 PY)	Complete f/u (per 100 PY)	Ratio
Placebo	3.1	0.9	3
2.5 mg Riva	6.3	1.4	4.5
5 mg Riva	9	1.8	5

In patients with complete follow-up, MI rates were 2 to 3-fold ↑, and mortality rates 5-fold ↑ in patients who experienced bleeding vs. those who did not

- ↑ bleeding led to both subject withdrawals and to ↑ MACE
- Differential dropout and informative censoring biases results in favor of Riva

FDA Briefing Document 2012

Do Missing Data Have a Material Impact on the ATLAS ACS-2 Results?

- Loss to follow-up rate exceeds the outcome event rate: Yes
- Missing data < 5% (or >20%): No (No)
- Missing data differential by treatment group and related to Rx: Yes
- Reasonable sensitivity analyses yield different results: Yes

Totality of evidence suggests the potential for missing data in ATLAS to have a material impact on trial interpretation

Krantz M and Kaul S, JACC 2013

Handling Missing Data in Clinical Trials
National Research Council, 2010

The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL REPORT

The Prevention and Treatment of Missing Data in Clinical Trials

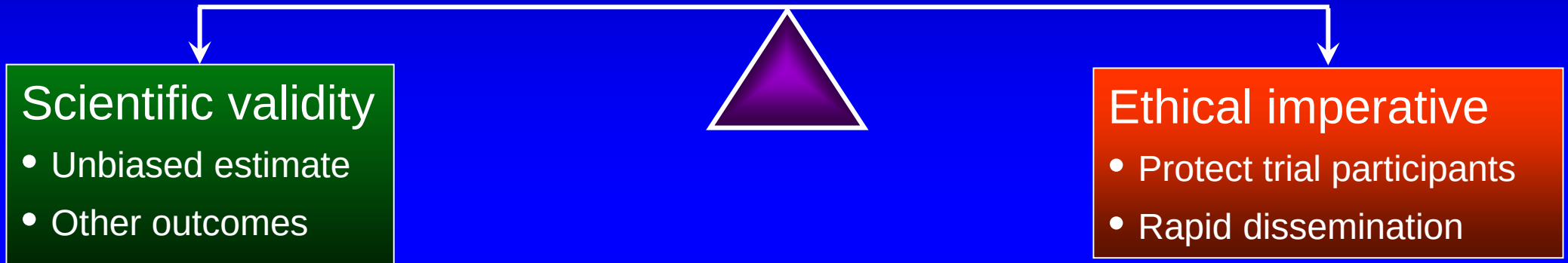
Roderick J. Little, Ph.D., Ralph D'Agostino, Ph.D., Michael L. Cohen, Ph.D., Kay Dickersin, Ph.D., Scott S. Emerson, M.D., Ph.D., John T. Farrar, M.D., Ph.D., Constantine Frangakis, Ph.D., Joseph W. Hogan, Sc.D., Geert Molenberghs, Ph.D., Susan A. Murphy, Ph.D., James D. Neaton, Ph.D., Andrea Rotnitzky, Ph.D., Daniel Scharfstein, Sc.D., Weichung J. Shih, Ph.D., Jay P. Siegel, M.D., and Hal Stern, Ph.D.

'Prevention' of missing data rather than 'treatment' remains the optimal approach to limit the problem and consequently enhance the credibility of causal inferences from clinical trials

Little et al, NEJM 2012;367:1355-50

Randomized Clinical Trial Stopped Early

Balancing Contrasting Goals



"Overly sanguine estimates of treatment effect result in misleading risk-benefit ratios, misguided practice recommendations, and suboptimal clinical practice"

Montori et al, *JAMA 2005;* 294:2203–2209

Premature Stopping of Trials
Recent Examples

- Trials stopped with prespecified stopping rules

  SPRINT, PARADIGM-HF, JUPITER

- Trials stopped without prespecified stopping rules

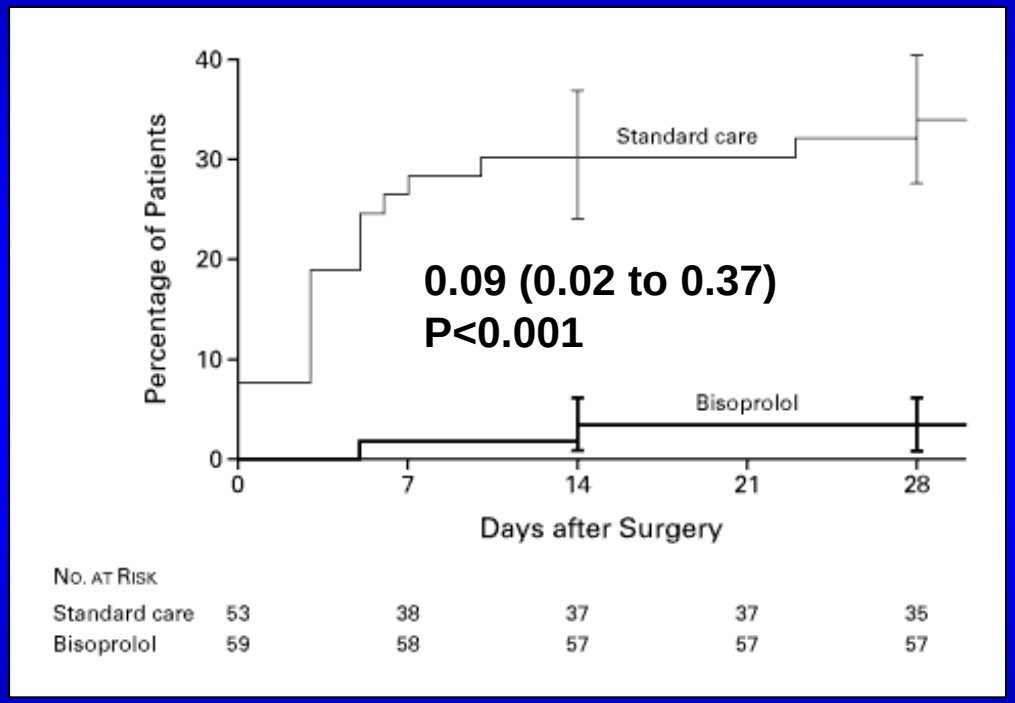
  - FAME-II (FFR in PCI; enrollment and follow-up truncated!)
  - PRAMI (PCI of non-infarct artery)

- Trials generally stopped for:

  - Unacceptable safety (“primum non nocere”)
  - Futility (ACCELERATE; CETP inhibitor evacetrapib)
  - Overwhelming benefit
    - Primary endpoint (SPRINT, JUPITER)
    - Primary endpoint and/or mortality (PARADGM-HF)
  - New external information of unequivocal efficacy & safety

Perioperative Beta-blockade (POBB) in High-Risk Patients Undergoing Vascular Surgery

Perils of Early Stopping

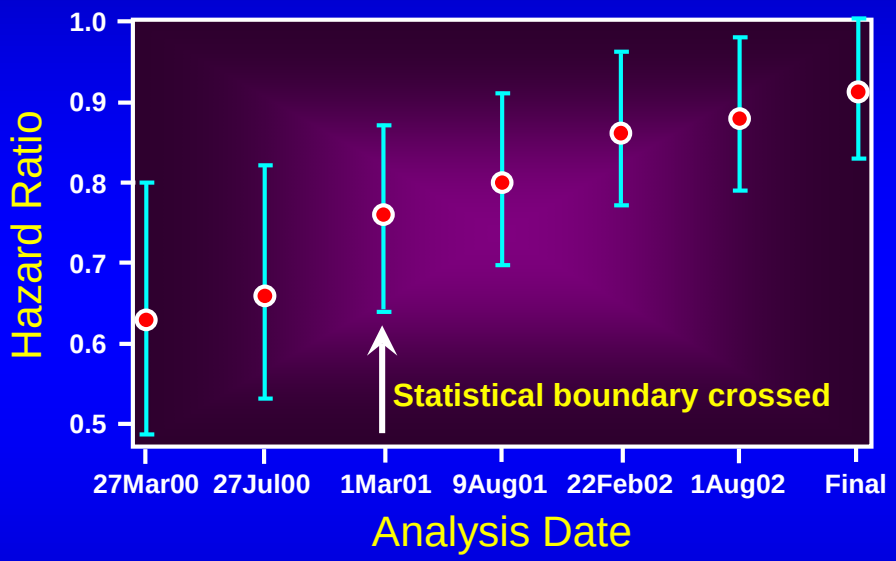


- Stopping rule
  - O'Brein-Fleming (P=0.001)
- 1st interim analysis at 112 pts
  - RR 0.09 (0.02 to 0.37)
  - P<0.001
- Planned to recruit 266 pts
- Expected RRR = 50%
- Implausible treatment effect
- Widely disseminated
- Changed practice guidelines
- Performance measure
- Never been replicated
- Recent meta-analyses show harm

Poldermans et al. NEJM 1999;341:1789-1794

Monitoring for Benefit in CHARM Trial

Perils of Early Stopping



- Stopping rule
  - Haybittle-Peto (P<0.001)
- 4th interim analysis
  - OR = 0.76 (0.64-0.87)
  - Logrank P=0.0006
- Final analysis
  - OR = 0.91 (0.83-1.00)
  - P = 0.055

"Random high" within the 1st year ("too good to be true")
"Regression to the truth" beyond 1st year

Pocock et al, *Clinical Trials* 2006; **3**: 513–521

Standards for Developing Trustworthy Clinical Practice Guidelines: IOM Report

- Establishing Evidence Foundations (Standard 5)
  - *Quality of evidence*
  - *Quantity of evidence (magnitude and precision)*
  - Consistency of aggregate available evidence
  - *Clear description of benefits and harms*
  - Rating of strength of recommendation

http://www.iom.edu/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx

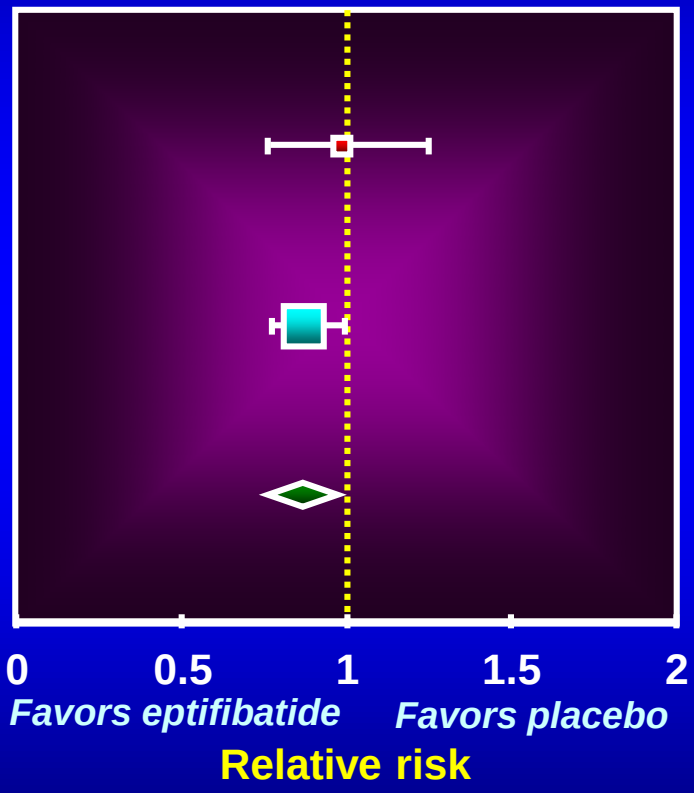
ACC/AHA Guideline Recommendations
Glycoprotein IIb/IIIa Inhibitor for UA/NSTEMI

	Class I (Benefit >>> risk) (*Highly recommended*)	Class II		Class III (Risk / No Benefit) (*Not recommended*)
		IIa (Benefit >>risk) (*Reasonably recommended*)	IIb (Benefit ≥ risk) (*May be considered*)	
Level A (Multiple randomized clinical trials)	High risk patients oriented to early invasive strategy (before or at PCI) Eptifibatide or tirofiban preferred			Abciximab in whom PCI not planned
Level B (Single randomized trial or nonrandomized studies)			"Upstream" use in high-risk pts (↑Tn, DM, ST↓) on ASA & clopidogrel and at low risk for bleeding	Low TIMI risk score or high bleeding risk and who are on ASA, Clopidogrel
Level C (Consensus opinion, case studies, or standard of care)		Recurrent ischemia during early conservative Rx with ASA, UFH, and clopidogrel		

Wright RS, Anderson JL et al. 2011; Jneid H et al, JACC/Circulation 2012

Primary Composite Endpoint in PURSUIT
Death or Nonfatal MI

Endpoint (first events*)	Eptifibatide (N=4722)	Placebo (N=4739)	Prevalence (%)	Risk ratio (95% CI)	ARD (%)
Death	165	175	16	0.95 (0.77-1.17)	0.20
Nonfatal MI	505	569	84	0.89 (0.80-1.00)	1.30
Composite	671 (15.7%)	744 (14.2%)		0.91 (0.82-1.00)	1.50

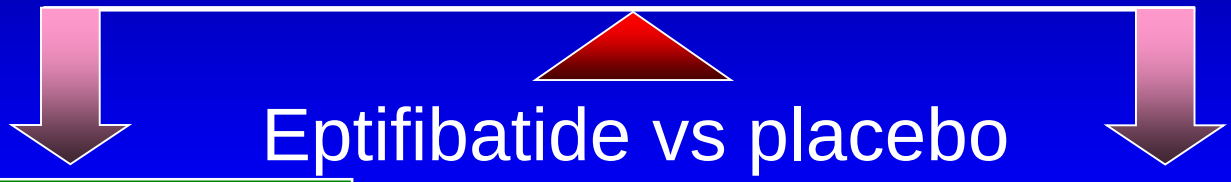


* Analysis based on time to first event (no double counting)

Benefit driven by nonfatal MI, primarily defined by biomarker elevation

Benefit-Risk Balance in PURSUIT

1000 Patients Treated with Eptifibatide Instead of Placebo



Benefit

- 15 ischemic endpoints prevented
  - 2 deaths
  - 13 nonfatal MIs

Risk

- 70 excess TIMI Major/Minor bleeds
  - 27 nonfatal TIMI Major bleeds
  - 43 nonfatal TIMI Minor bleeds

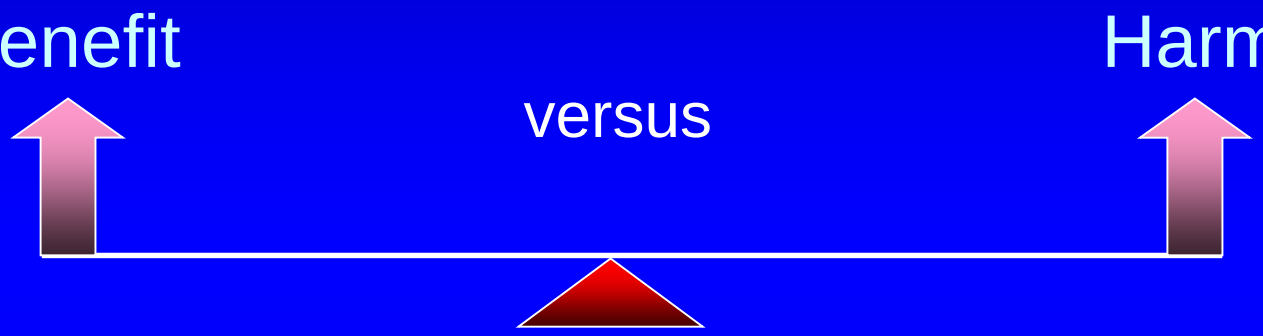
or

- 32 excess GUSTO bleeds
  - 8 nonfatal severe bleeds
  - 24 nonfatal moderate bleeds

Does the evidence favor Class I (benefit >>> risk) recommendation for eptifibatide in ACS?

Quantification of Net Clinical Benefit

Balance Sheet of NNT and NNH (PURSUIT)



NNT for death/nonfatal MI = 67
NNT for nonfatal MI = 77

NNH for TIMI major bleed = 37
NNH for GUSTO mod/severe bleed = 31

Desirable benefit-risk = NNT << NNH or NNT/NNH <1 (assuming benefits and harms are equal)

Consumer Reports Guide
Interpreting 'Positive' (P<α) Trials

Rank	Quality	Quantity	Benefit-Risk
★★★★★	High	• Large effect size • Statistically persuasive	B>>>R
★★★★	High	• Modest effect size • Statistically persuasive	B>>R
★★★	Modest	• Modest effect size • Statistically persuasive	B>R
★★	Modest	• Small effect size • Statistically not persuasive	B>R
★	Low	• Small effect size • Statistically not persuasive	B=/<R

Impactful RCTs of Last Two Years

- PARADIGM-HF (AHA/NEJM 2014)
  - Sacubitril + Valsartan vs. Valsartan + Placebo in CHF (NYHA Class ≥II + EF≤40%)
- IMPROVE-IT (AHA 2014, NEJM 2015)
  - Ezetimibe + Simvastatin vs. Simvastatin + Placebo post-Acute Coronary Syndrome
- EMPA-REG OUTCOME (EASD/NEJM 2015)
  - Empagliflozin (SGLT2 inhibitor) vs. Placebo in Type 2 Diabetes
- SPRINT (AHA/NEJM 2015)
  - Intensive vs. Standard BP control in patients with SBP>130mmHg + increased CV risk

Impactful RCTs of Last Two Years
Key Quality Attributes

Trial	Type	Blind	Power (1-β)	MDD (δ)	Missing data	Prematurely stopped
PARADIGM-HF (N=8,442)	Superiority	DB	80%	RR 0.85	0.2%	Yes*
IMPROVE-IT (N=18,144)	Superiority	DB	90%	RR 0.91	11%	No
EMPA-REG OUTCOME (N=7,042)	Noninferiority (NI) Superiority (S)	DB	90% 80%	HR 1.30 (NI) HR 0.785 (S)	3%	No
SPRINT (N=9,361)	Superiority	OL	88.7%	RR 0.80	5.5%**	Yes***

OL=open label; DB=double blind; MDD=minimal detectable difference; β=type II error
* based on PEP & mortality (overwhelming benefit: P=0.001 @ 3rd interim analysis)
**lost to follow up (N=245), withdrew consent (N=275); ? impact of missing data on outcomes explored
*** not clear if based on overwhelming benefit in PEP alone or mortality as well

Impactful RCTs of Last Two Years

Quantity of Evidence

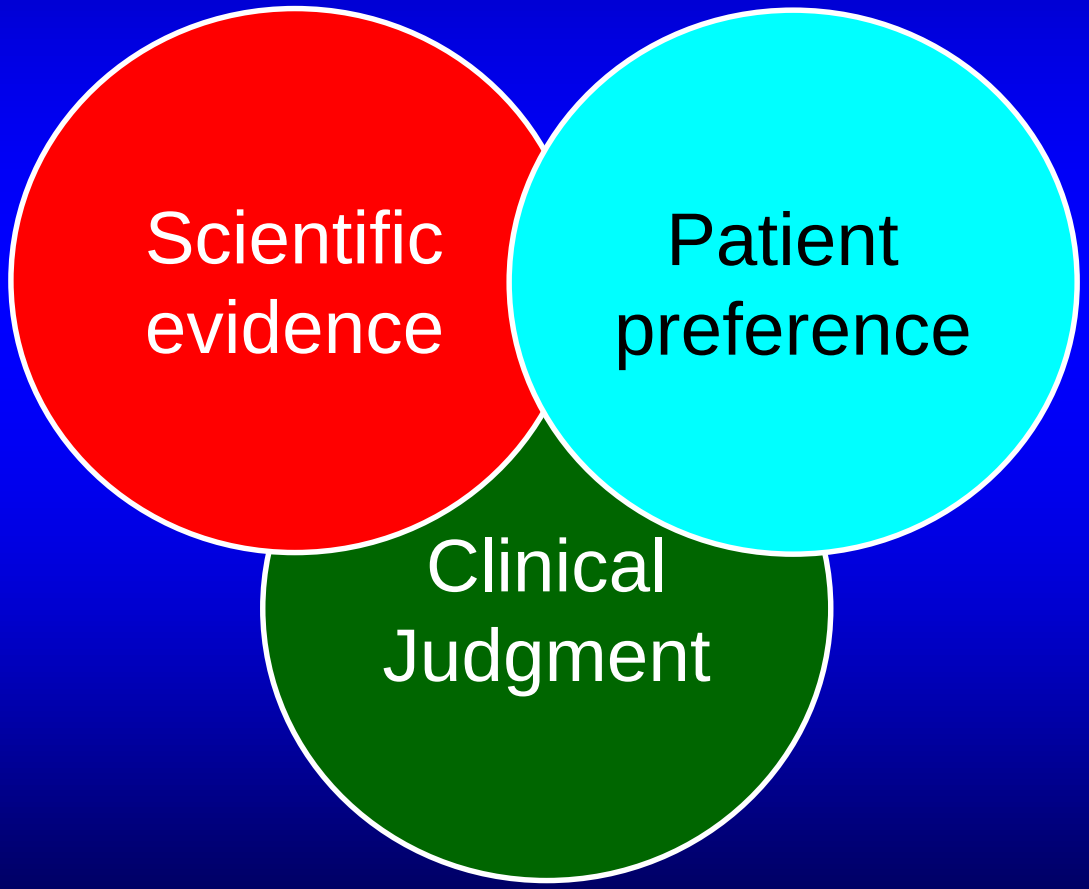
Trial	Endpoint	Outcome			Substantial evidence
		RR (95% CI)	P value	Min. Bayes factor	
PARADIGM-HF (N=8,442)	CVD, HF hosp.	0.80 (0.73, 0.87)	<0.0001	<0.0001	Yes
	CVD	0.80 (0.71, 0.89)	<0.0001	0.0003	Yes
	ACM	0.84 (0.76, 0.93)	<0.001	0.0035	Yes
IMPROVE-IT (N=18,144)	CVD, MI, Stroke	0.94 (0.89, 0.99)	0.016	0.06	No
EMPA-REG OUTCOME (N=7,042)	CVD, MI, Stroke	0.86 (0.74, 0.99)	0.038	0.131	No
	CVD	0.62 (0.49, 0.77)	0.0001	0.0004	Yes
	ACM	0.68 (0.57, 0.82)	0.0001	0.0006	Yes
SPRINT (N=9,361)	CVD, MI/ACS, Stroke, HF	0.75 (0.64, 0.89)	<0.001	0.004	Yes
	CVD	0.57 (0.38, 0.85)	0.005	0.021	Yes
	ACM	0.73 (0.60, 0.90)	0.003	0.014	Yes

Consumer Reports Guide
Interpreting 'Positive' (P<α) Trials

Rank	Quality	Quantity	Benefit-Risk	Trial
★★★★★	High	• Large effect size • Statistically persuasive	B>>>R	- PARADIGM-HF - EMPA-REG (mortality)
★★★★	High	• Modest effect size • Statistically persuasive	B>>R	
★★★	Modest	• Modest effect size • Statistically persuasive	B>R	SPRINT
★★	Modest	• Small effect size • Statistically not persuasive	B>R	IMPROVE-IT
★	Low	• Small effect size • Statistically not persuasive	B=/<R	(PURSUIT)

“Evidence-Based” Not “Evidence-Bound”

Three Key Dimensions



“There are no facts, only interpretations”

Friedrich Nietzsche

Consumer Reports Guide to Interpreting Trials
Conclusions

- Statistical significance vs clinical importance
  - Develop validated, domain-specific thresholds for clinical importance
  - Apply during trial design & interpretation
  - Evidence appraisal (regulatory approval, reimbursement, guidelines)
- Missing data
  - Potential to invalidate ITT analysis (best to ‘prevent’, no good ‘treatment’)
  - Use multiple imputation methods/worst case scenario to assess impact
- Premature truncation
  - Pre-specify stopping rules
  - Unacceptable safety, futility and overwhelming benefit (preferably mortality)
- Composite endpoints
  - Avoid clinically unimportant and unvalidated outcomes
  - Assess for large treatment gradients (heterogeneity of treatment effect)

Consumer Reports Guide to Interpreting Trials
Conclusions

- Noninferiority
  - Margin should be based on clinical judgment and statistical reasoning
  - New Rx should offer tangible ancillary benefits (safety, cost, convenience)
- Subgroup analysis
  - Prespecified based on biological plausibility or prior evidence; pre-randomization
  - Test for interaction
  - Minimize multiple subgroups, plan ahead and adjust for multiplicity
- Benefit-risk assessment
  - Qualitative science grounded in quantitative data & dependent on judgment
  - Effect size, seriousness of events, and availability of safer alternatives should drive benefit-risk tradeoffs
  - Incorporate patient's views of acceptable risk (values and preferences)
  - Evolve from the mindset to "ensure drug safety" to ensure "favorable benefit-risk profile"