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
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

- \( \text{vehemence} \propto \text{evidence}^{-1} \)

- \( \text{vehemence} \propto \text{eminence}^{2} \)

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
<table>
<thead>
<tr>
<th>Organization</th>
<th>Evidence Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDA/CDER</td>
<td>“Substantial evidence of Effectiveness”</td>
</tr>
<tr>
<td>FDA/CDRH (PMA)</td>
<td>“Reasonable assurance of Safety and Efficacy”</td>
</tr>
<tr>
<td>FDA/CDRH (510k)</td>
<td>“Substantial evidence of Equivalence”</td>
</tr>
<tr>
<td>ACC/AHA</td>
<td>“Useful and Effective”</td>
</tr>
<tr>
<td>CMS</td>
<td>“Reasonable and Necessary”</td>
</tr>
<tr>
<td>Payers</td>
<td>“Usual and Customary”</td>
</tr>
<tr>
<td>Courts</td>
<td>“Prudent and Cautious”</td>
</tr>
<tr>
<td>Consumer Reports</td>
<td>“Reliable and meaningful”</td>
</tr>
</tbody>
</table>
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

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

<table>
<thead>
<tr>
<th>CFR</th>
<th>Statutory criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FDC Act 1962</strong></td>
<td>“Substantial” evidence of effectiveness consisting of “adequate and well-controlled investigations”, i.e., two separate trials each with $p&lt;0.05$ ($0.05 \times 0.05 = 0.0025$ divided by 2 = 0.001)</td>
</tr>
<tr>
<td><strong>FDA Evidence Guidance for Industry, 1998</strong></td>
<td>“A highly persuasive statistical finding (a $p$ value &lt;0.001) in a single trial with some other indication of the study’s reliability (e.g., multicenter with no center driving the results)”</td>
</tr>
<tr>
<td><strong>FDAMA 115 (1998)</strong></td>
<td>“One adequate and well-controlled study and confirmatory evidence.”</td>
</tr>
</tbody>
</table>
Replication Probability and P-values

<table>
<thead>
<tr>
<th>P value of initial expt.</th>
<th>Probability of $p&lt;0.05$ when the first observed difference is true</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>.37</td>
</tr>
<tr>
<td>0.05</td>
<td>.50</td>
</tr>
<tr>
<td>0.03</td>
<td>.58</td>
</tr>
<tr>
<td>0.01</td>
<td>.73</td>
</tr>
<tr>
<td>0.005</td>
<td>.80</td>
</tr>
<tr>
<td>0.001</td>
<td>.91</td>
</tr>
</tbody>
</table>

‘Replication is at the heart of scientific endeavor’

**PURSUIT Trial (Death or MI)**

**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 ($\delta$) of 20% risk reduction.

**Conclusion:**
- Eptifibatide is superior to aspirin and heparin in NSTE ACS.

**Does the evidence justify strong recommendation?**

<table>
<thead>
<tr>
<th>OUTCOME</th>
<th>EVENT</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Ept</td>
<td>671</td>
<td>4051</td>
</tr>
<tr>
<td>Con</td>
<td>744</td>
<td>3995</td>
</tr>
<tr>
<td>Total</td>
<td>1415</td>
<td>8046</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Event Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eptifibatide</td>
<td>4722</td>
<td>14.2</td>
</tr>
<tr>
<td>Control</td>
<td>4739</td>
<td>15.7</td>
</tr>
</tbody>
</table>

**RRR** 9%  
95% CI 1%-18%  
P (2-tailed) 0.04
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/H₀)/Prob (Data/H₁) (likelihood ratio)
  - H₀ = Null hypothesis; H₁ = alternative hypothesis
  - Odds = Probability/(1-Probability)
  - Probability = Odds/(1+Odds)
  - Minimum BF =exp(-0.5z²)

Bayes factor is a comparison of how well two hypotheses predict the data: smaller the BF, stronger the evidence against H₀
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

<table>
<thead>
<tr>
<th>P Value (Z score)</th>
<th>Minimum Bayes Factor</th>
<th>Decrease in Probability of Null Hypothesis, %</th>
<th>Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.04 (2.03)</td>
<td>0.13 (1/7.7)</td>
<td>From 75 To 50 No Less Than 28</td>
<td>Moderate</td>
</tr>
<tr>
<td>P Value (Z score)</td>
<td>Minimum Bayes Factor</td>
<td>Decrease in Probability of Null Hypothesis, %</td>
<td>Strength of Evidence</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------</td>
<td>---------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>From</td>
<td>To No Less Than</td>
</tr>
<tr>
<td>0.10 (1.64)</td>
<td>0.26 (1/3.8)</td>
<td>75</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05 (1.96)</td>
<td>0.15 (1/6.8)</td>
<td>75</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.03 (2.17)</td>
<td>0.095 (1/11)</td>
<td>75</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01 (2.58)</td>
<td>0.036 (1/28)</td>
<td>75</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001 (3.28)</td>
<td>0.005 (1/216)</td>
<td>75</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **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
**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
“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

Risk Ratio (95% CI)

0.85 1.0

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**

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

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

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

Rivaroxaban (both doses)

HR 0.84 (0.74-0.96)
mITT p = 0.008
ITT p = 0.002
NNT = 59

Placebo

No. at Risk
Placebo 5113 4307 3470 2664 1831 1079 421
Riva 10229 8502 6753 5137 3554 2084 831

Months After Randomization

2 Yr KM Estimate
10.7%
8.9%

Mega et al, NEJM 2012;366:9-19
## Missing Data in Contemporary ACS Trials

### ATLAS ACS-2 TIMI 51

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

- 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**

<table>
<thead>
<tr>
<th>Group</th>
<th>Stratum</th>
<th>Rivaroxaban</th>
<th>Placebo</th>
<th>Relative Risk (95% CI)</th>
<th>Excess # event , Riva</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pooled</strong></td>
<td>All</td>
<td>626/10229 (6.1)</td>
<td>376/5113 (7.4)</td>
<td>0.84 (0.74-0.96)</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>666/10229 (6.5)</td>
<td>376/5113 (7.4)</td>
<td>0.89 (0.78-1.00)</td>
<td></td>
</tr>
<tr>
<td>2 (+ DAPT)</td>
<td>All</td>
<td>575/9532 (6.0)</td>
<td>340/4760 (7.1)</td>
<td>0.86 (0.75-0.98)</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>598/9532 (6.3)</td>
<td>340/4760 (7.1)</td>
<td>0.88 (0.77-1.00)</td>
<td></td>
</tr>
<tr>
<td><strong>2.5 mg</strong></td>
<td>All</td>
<td>313/5114 (6.1)</td>
<td>376/5113 (7.4)</td>
<td>0.84 (0.72-0.97)</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>326/10229 (6.4)</td>
<td>376/5113 (7.4)</td>
<td>0.87 (0.75-1.00)</td>
<td></td>
</tr>
<tr>
<td>2 (+ DAPT)</td>
<td>All</td>
<td>286/4765(6.0)</td>
<td>340/4760 (7.1)</td>
<td>0.85 (0.72-0.99)</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>293/4765 (6.2)</td>
<td>340/4760 (7.1)</td>
<td>0.86 (0.74-1.00)</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incomplete follow up for Rivaroxaban:</td>
<td>2192</td>
</tr>
<tr>
<td>Excess Rivaroxaban events to p &gt; 0.05:</td>
<td>7 to 40</td>
</tr>
<tr>
<td>% Patients with TIMI major/minor bleeds who experience MACE:</td>
<td>37%</td>
</tr>
<tr>
<td>No. of TIMI major minor bleeds to yield 7-40 excess MACE (7/0.37 to 40/0.37)</td>
<td>19 to 108</td>
</tr>
<tr>
<td>No. of Rivaroxaban incomplete f/u, no MACE, but TIMI major/minor bleed</td>
<td>98</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th>Group</th>
<th>Incomplete f/u (per 100 PY)</th>
<th>Complete f/u (per 100 PY)</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>3.1</td>
<td>0.9</td>
<td>3</td>
</tr>
<tr>
<td>2.5 mg Riva</td>
<td>6.3</td>
<td>1.4</td>
<td>4.5</td>
</tr>
<tr>
<td>5 mg Riva</td>
<td>9</td>
<td>1.8</td>
<td>5</td>
</tr>
</tbody>
</table>

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
‘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

Scientific validity
- Unbiased estimate
- Other outcomes
- Benefit-risk ratio

Ethical imperative
- Protect trial participants
- Rapid dissemination

"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 (PARADIGM-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
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

## ACC/AHA Guideline Recommendations

### Glycoprotein IIb/IIIa Inhibitor for UA/NSTEMI

<table>
<thead>
<tr>
<th>Class I</th>
<th>Class II</th>
<th>Class III</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Benefit &gt;&gt;&gt; risk)</td>
<td>Ilia (Benefit &gt;&gt; risk)</td>
<td>(Risk / No Benefit)</td>
</tr>
<tr>
<td>(Highly recommended)</td>
<td>(Reasonably recommended)</td>
<td>(Not recommended)</td>
</tr>
</tbody>
</table>

### Level A
- High risk patients oriented to early invasive strategy (before or at PCI)
- Eptifibatide or tirofiban preferred
- Abciximab in whom PCI not planned

### Level B
- “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
- 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**

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

* 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

Eptifibatide vs placebo

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)

Benefit versus Harm

Eptifibatide

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<\(\alpha\)) Trials

<table>
<thead>
<tr>
<th>Rank</th>
<th>Quality</th>
<th>Quantity</th>
<th>Benefit-Risk</th>
</tr>
</thead>
</table>
| ★★★★★ | 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

<table>
<thead>
<tr>
<th>Trial</th>
<th>Type</th>
<th>Blind</th>
<th>Power $(1-\beta)$</th>
<th>MDD $(\delta)$</th>
<th>Missing data</th>
<th>Prematurely stopped</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARADIGM-HF</td>
<td>Superiority</td>
<td>DB</td>
<td>80%</td>
<td>RR 0.85</td>
<td>0.2%</td>
<td>Yes*</td>
</tr>
<tr>
<td>(N=8,442)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMPROVE-IT</td>
<td>Superiority</td>
<td>DB</td>
<td>90%</td>
<td>RR 0.91</td>
<td>11%</td>
<td>No</td>
</tr>
<tr>
<td>(N=18,144)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMPA-REG OUTCOME</td>
<td>Noninferiority (NI)</td>
<td>DB</td>
<td>90% 80%</td>
<td>HR 1.30 (NI) HR 0.785 (S)</td>
<td>3%</td>
<td>No</td>
</tr>
<tr>
<td>(N=7,042)</td>
<td>Superiority (S)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPRINT</td>
<td>Superiority</td>
<td>OL</td>
<td>88.7%</td>
<td>RR 0.80</td>
<td>5.5%**</td>
<td>Yes***</td>
</tr>
<tr>
<td>(N=9,361)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OL=open label; DB=double blind; MDD=minimal detectable difference; $\beta$=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

<table>
<thead>
<tr>
<th>Trial</th>
<th>Endpoint</th>
<th>Outcome</th>
<th>RR (95% CI)</th>
<th>P value</th>
<th>Min. Bayes factor</th>
<th>Substantial evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PARADIGM-HF</strong> (N=8,442)</td>
<td>CVD, HF hosp.</td>
<td></td>
<td>0.80 (0.73, 0.87)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>CVD</td>
<td></td>
<td>0.80 (0.71, 0.89)</td>
<td>&lt;0.0001</td>
<td>0.0003</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>ACM</td>
<td></td>
<td>0.84 (0.76, 0.93)</td>
<td>&lt;0.001</td>
<td>0.00035</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>IMPROVE-IT</strong> (N=18,144)</td>
<td>CVD, MI, Stroke</td>
<td></td>
<td>0.94 (0.89, 0.99)</td>
<td>0.016</td>
<td>0.06</td>
<td>No</td>
</tr>
<tr>
<td><strong>EMPA-REG OUTCOME</strong> (N=7,042)</td>
<td>CVD, MI, Stroke</td>
<td></td>
<td>0.86 (0.74, 0.99)</td>
<td>0.038</td>
<td>0.131</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>CVD</td>
<td></td>
<td>0.62 (0.49, 0.77)</td>
<td>0.0001</td>
<td>0.0004</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>ACM</td>
<td></td>
<td>0.68 (0.57, 0.82)</td>
<td>0.0001</td>
<td>0.0006</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>SPRINT</strong> (N=9,361)</td>
<td>CVD, MI/ACS, Stroke, HF</td>
<td></td>
<td>0.75 (0.64, 0.89)</td>
<td>&lt;0.001</td>
<td>0.004</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>CVD</td>
<td></td>
<td>0.57 (0.38, 0.85)</td>
<td>0.005</td>
<td>0.021</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>ACM</td>
<td></td>
<td>0.73 (0.60, 0.90)</td>
<td>0.003</td>
<td>0.014</td>
<td>Yes</td>
</tr>
</tbody>
</table>
## Consumer Reports Guide
### Interpreting ‘Positive’ (P<\(\alpha\)) Trials

<table>
<thead>
<tr>
<th>Rank</th>
<th>Quality</th>
<th>Quantity</th>
<th>Benefit-Risk</th>
<th>Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>★★★★★</td>
<td>High</td>
<td>Large effect size Statistically persuasive</td>
<td>B&gt;&gt;&gt;R</td>
<td>PARADIGM-HF, EMPA-REG (mortality)</td>
</tr>
<tr>
<td>★★★★★</td>
<td>High</td>
<td>Modest effect size Statistically persuasive</td>
<td>B&gt;&gt;R</td>
<td></td>
</tr>
<tr>
<td>★★★★</td>
<td>Modest</td>
<td>Modest effect size Statistically persuasive</td>
<td>B&gt;R</td>
<td>SPRINT</td>
</tr>
<tr>
<td>★★★</td>
<td>Modest</td>
<td>Small effect size Statistically not persuasive</td>
<td>B&gt;R</td>
<td>IMPROVE-IT</td>
</tr>
<tr>
<td>★★</td>
<td>Modest</td>
<td>Small effect size Statistically not persuasive</td>
<td>B=/&lt;R</td>
<td>(PURSUIT)</td>
</tr>
<tr>
<td>★</td>
<td>Low</td>
<td>Small effect size Statistically not persuasive</td>
<td>B=/&lt;R</td>
<td></td>
</tr>
</tbody>
</table>
“Evidence-Based” Not “Evidence-Bound”
Three Key Dimensions

Scientific evidence

Patient preference

Clinical Judgment
“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)
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”