Some Current Controversies in Clinical Trials Research

Stuart Pocock

London School of Hygiene and Tropical Medicine
SUNY at Buffalo in the 1970s under Marvin’s Leadership

definitive developments in cancer clinical trials

Cooperative Groups (eg ECOG)

statistics, data management, computing, collaboration

Statistical Developments

survival analysis, randomization, stopping guidelines

People

Colin Begg, Michael Feldstein, Rich Gelber, Jim Hanley, Jack Kalbfleisch, Phil Lavin, Steve Lagakos, Marcello Pagano, Greg Pavlov, Ross Prentice, Dave Schoenfeld, Norm Severo, Ken Stanley, Paul Carbone, Thelma Zelen

Cancer Trials Meeting, Airlie House 1972
**Current Controversies**

When to Stop a Trial Early

Adaptive Designs

Subgroup Analyses

Personalized Medicine

Scare Stories

Non-inferiority Trials

Comparative Effectiveness

---

**When to Stop a Trial Early**

Statistical stopping “rules” are guidelines only

an objective aid to DSMB wise judgment

often guidelines for primary endpoint only

**stopping for efficacy, harm or futility**

DSMB acts on totality of evidence

Stopping for efficacy:
need proof beyond reasonable doubt (eg. P<.001)

stopping early [✓] change in future practice
PRAMI trial: Preventive Angioplasty in Myocardial Infarction
[NEJM Sept 2013]

treat culprit lesion only OR other narrowed arteries as well

trial stopped early (mean 23 months follow-up)

Preventive Angioplasty

<table>
<thead>
<tr>
<th></th>
<th>NO (N=231)</th>
<th>YES (N=234)</th>
<th>hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>primary endpoint</td>
<td>53</td>
<td>21</td>
<td>0.35(0.21,0.58)</td>
</tr>
<tr>
<td>refractory angina</td>
<td>30</td>
<td>12</td>
<td>0.35(0.18,0.69)</td>
</tr>
<tr>
<td>nonfatal MI</td>
<td>20</td>
<td>7</td>
<td>0.32(0.13,0.75)</td>
</tr>
<tr>
<td>cardiac death</td>
<td>10</td>
<td>4</td>
<td>0.34(0.11,1.08)</td>
</tr>
</tbody>
</table>

Issues to consider

- a huge treatment difference: too good to be true?
- trial stopped early: tendency to exaggerate efficacy
- smallish trial with rather few events
- trial not blinded, potential for bias
- "hypothesis generating", rather than changing practice?
- another larger trial needed
**SHIFT trial**  
[Lancet online Aug 29, 2010]

**ivabradine vs placebo** in 6558 heart failure patients

**second planned interim efficacy analysis June 2009**
primary endpoint:
CV death and CHF hosp\(^\text{a}\) hazard ratio 0.77 P<.0001
all cause death hazard ratio 0.77 P=.0014

DMC decided **not** to recommend stopping early:
previous BEAUTIFUL trial “negative”, subgroup issues,
only a few months to go, event validation issues

**Final Results:**
primary endpoint  hazard ratio 0.82 P<.0001
all cause death  hazard ratio 0.90 P=.092

---

**Stopping for Harm: the agonizing negative trend**

**more frequent looks, more lenient stopping boundary**

**ILLUMINATE trial**  
[NEJM 2007; 357 p2109-]
torcetrapib vs. placebo in 15067 high risk patients

primary endpoint: CHD death, MI, stroke + unstable angina
torcetrapib raises HDL cholesterol (but also raises BP)
emerging evidence of harm: monthly report 30 Nov 2006
82 vs 51 deaths P=0.007 (stopping guideline P<0.01)
DSMB teleconference 1 Dec 2006, recommendation to stop
Sponsor stopped all torcetrapib trials on 2 Dec 2006

---
Types of Adaptive Design

After unblinded interim analysis:

Increase sample size

Drop treatment arms/doses

Change entry criteria

Change randomization ratio

Change primary endpoint

need to preserve statistical rigour and trial’s integrity

Adaptive Options at a Planned Interim Analysis
using conditional power (CP)

Interim Difference       Decision
→ Amazing                Stop now for efficacy
Favorable                Continue to original N, CP(N) > 90%
✗ Promising              Extend to new N* so that CP(N*)=90%(say)
✗ Not brilliant           Extend to Nmax so that P(Nmax)>30%(say)
Disappointing            Continue to original N
→ Hopeless               Stop now for futility or harm
→ conventional data monitoring boundaries (optional)
✗ adaptive options
CHAMPION PCI trial used adaptive design
cangrelor vs clopidogrel in 8750 ACS patients
primary endpoint: death, MI, revasc. at 48 hrs
interim analysis after 70% evaluated
5 options: stop early for efficacy
  proceed to N = 8750 if ≥ 80% conditional power
  expand up to N = 15000 to have 80% CP
  proceed to N = 8750 if less promising
  stop early for futility if < 20% CP
detailed adaptive charter approved by FDA

CHAMPION PCI Results [NEJM Nov 2009]
trial stopped early for futility

<table>
<thead>
<tr>
<th></th>
<th>cangrelor</th>
<th>clopidogrel</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>3889</td>
<td>3865</td>
</tr>
<tr>
<td>primary endpoint at 48 hours</td>
<td>290</td>
<td>276</td>
</tr>
<tr>
<td>P</td>
<td>.59</td>
<td></td>
</tr>
</tbody>
</table>

at most 1 in 4 adaptive designs will actually adapt
Statistical Penalty for Sample Size Re-estimation?

Chen et al [Stats in Med 2004; 23 p 1023]

"Increasing sample size when unblinded interim result is promising will not inflate type I error. No statistical adjustment required"

Mehta and Pocock [Stats in Med 2011; 30 p 3267-]

adaptive increase when in the promising one

The Traumas of Subgroup Analysis

The CHARISMA trial [NEJM 12 March 2006]
15,603 high risk patients on low dose aspirin
primary endpoint: MI, stroke or CV death

clopidogrel  placebo

<table>
<thead>
<tr>
<th>Overall</th>
<th>6.8%</th>
<th>7.3%</th>
<th>P=.22</th>
</tr>
</thead>
<tbody>
<tr>
<td>symptomatic (N=12153)</td>
<td>6.9%</td>
<td>7.9%</td>
<td>P=.046</td>
</tr>
<tr>
<td>asymptomatic (N=3284)</td>
<td>6.6%</td>
<td>5.5%</td>
<td>P=.2</td>
</tr>
</tbody>
</table>

interaction test P=.045
plausible, provocative, special pleading

editorial: “the charisma of extracting favourable subgroups should be resisted".
Results by Geographic Region: tricky to Interpret

PLATO trial: ticagrelor vs clopidogrel in ACS

Primary Endpoint by Region

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia/Australia</td>
<td>1714</td>
<td>0.80 (0.61, 1.04)</td>
</tr>
<tr>
<td>Central/South America</td>
<td>1237</td>
<td>0.86 (0.65, 1.13)</td>
</tr>
<tr>
<td>Europe/Middle East/Africa</td>
<td>13859</td>
<td>0.80 (0.72, 0.90)</td>
</tr>
<tr>
<td>North America</td>
<td>1814</td>
<td>1.25 (0.93, 1.67)</td>
</tr>
</tbody>
</table>

interaction P=0.05

ticagrelor is superior, unless you're American!?

One of 32 subgroup analyses, beware of data dredging

Primary Endpoint for US and non-US [post hoc interaction P=.009]
median aspirin dose US vs non-US

≥ 300mg dose in 54% US and 1.7% non-US

Ticagrelor vs Clopidogrel by Aspirin Dose

[interaction P=.0006]

ASA low (<300mg): HR (95% CI), 0.79 (0.71, 0.88)
ASA high (≥300mg): HR (95% CI), 1.45 (1.01, 2.09)
Questions?

Is the US anomaly a reality, or just a chance finding?

Is maintenance aspirin dose an explanation, or a consequence of data dredging?

Is the FDA boxed warning “use of ticagrelor with aspirin doses exceeding 100 mg/day decreases its effectiveness” a wise decision?

Value of stratifying patients into risk groups
[NEJM 2008:358 p2076]

Invasive vs Conservative Strategy in ACS

IPD Meta-analysis [JACC 2010:55 p 2435-]

CV death or MI, hazard ratio 0.80 P=.005

<table>
<thead>
<tr>
<th>risk category</th>
<th>N</th>
<th>invasive</th>
<th>conservative</th>
<th>difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>1822</td>
<td>7.3%</td>
<td>8.4%</td>
<td>-1.1%</td>
</tr>
<tr>
<td>moderate</td>
<td>1806</td>
<td>10.5%</td>
<td>12.7%</td>
<td>-2.2%</td>
</tr>
<tr>
<td>high</td>
<td>918</td>
<td>19.3%</td>
<td>23.5%</td>
<td>-4.2%</td>
</tr>
<tr>
<td>very high</td>
<td>921</td>
<td>29.4%</td>
<td>41.4%</td>
<td>-12.0%</td>
</tr>
</tbody>
</table>

invasive strategy more appropriate in high risk patients
Personalised risk-benefit analysis

eg prasugrel vs clopidogrel in ACS patients
[Circ Cardiovas Qual Outcomes 2013]

TRITON TIMI 38: ACS patients undergoing PCI

overall prasugrel reduces ischaemia but increases bleeding

Major Ischaemia Risk Model
14 predictors of which 4 interact with treatment

Bleeding Risk Model
15 predictors of which 5 interact with treatment

Individual Patient’s Net Predicted Risk for Ischemia and Bleeding

42% of patients predicted to benefit from prasugrel
Statistical Cautions needed re
the Allure of Personalised (Stratified) Medicine

Stratify on new biomarker or genotype

biomarker enhances risk prediction?
given known predictors, how great is the gain?

biomarker enables targeted treatment?
therapy effective only in biomarker positive patients?

multiplicity: many biomarkers studied
risk of false positives

choice of cut-off: arbitrary?
over-simplifies a continuum

More sensible handling of drug safety issues

Scare stories, politics and the media

eg Avandia and myocardial infarction

activists ↔ defensive companies

 objective
unbiased
evidence

how can we avoid over-reaction
what's the real evidence
what's the appropriate consequences

wise decisions by regulatory authorities etc
Need wiser Interpretation of Weak Evidence

Rosiglitazone (Avandia) and cardiovascular risk

Meta-analysis of 42 trials  [Nissen & Wolski NEJM 2007]

Rosi vs Control  odds ratio (95% CI)
myocardial infarction  1.43 (1.03, 1.98)  P=.04
cardiovascular death  1.64 (0.98, 2.74)  P=.06

weak evidence of potential harm
mostly small trials, unvalidated events
high profile, Congress involved, FDA under attack

“I was truly frightened on behalf of our patients”
The London Times (business section)

“Alarmist headlines and confident declarations help nobody”
The Lancet

“Meta-analysis seems a rushed and incomplete examination”
Nature
RECORD Trial  [Lancet 2009; 373 p 2125-]

A large pragmatic cardiovascular safety trial requested by the EMEA

Rosi + M or S vs Metformin + Sulfonylurea

4458 diabetic patients, mean 5.5 years follow-up
primary endpoint: any CV hospitalization or death
non inferiority trial
powered to rule out 20% excess risk on rosi
RECORD was not double blind, potential bias?

RECORD trial Pre-planned Primary Outcome

Cardiovascular Deaths and Hospitalisations unaffected

[Graph showing cumulative percentage over time (years) with HR 0.99 (95% CI 0.85-1.16)]
So, Rosiglitazone is largely history

risk of myocardial infarction exaggerated, inconclusive

strong evidence re heart failure, bone fractures and weight gain

EMA:
"benefits no longer outweigh the risks: drug is suspended"

FDA:
panel re-assessment in June 2013
"rosi not associated with excess myocardial infarction"
FDA Guidance for Industry (Dec 2008)

Cardiovascular Risk in New Antidiabetic Therapies

Non-inferiority CV safety trial of new drug vs placebo

Primary endpoint: CV death, myocardial infarction, stroke

1) to get approval, need trial evidence
to rule out unacceptable (80%) excess CV risk

2) post-approval, need longer, larger trial to establish CVsafety more clearly (ie rule out 30% excess risk)

3) evidence of CV benefit would be a bonus

SAVOR-TIMI 53 trial [NEJM 2013; 369 p 1317-]

Saxagliptin vs Placebo in 16,492 high risk type II diabetics

788 sites in 26 countries, median 2.1 years follow-up

<table>
<thead>
<tr>
<th></th>
<th>saxagliptin [N=8280]</th>
<th>placebo [N=8212]</th>
<th>hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>primary endpoint</td>
<td>613</td>
<td>609</td>
<td>1.00 (0.89 to 1.12)</td>
</tr>
<tr>
<td>(CV death, MI, stroke)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>heart failure hosp&quot;</td>
<td>289</td>
<td>228</td>
<td>1.27 (1.07 to 1.51)</td>
</tr>
</tbody>
</table>

\[P=.007\]

primary endpoint: non-inferiority established, but no benefit

heart failure: given multiple testing, a false positive?
OPTIMIZE trial [JAMA online 31 Oct 2013]

3 vs 12 months dual antiplatelet therapy after drug-eluting stent

**primary endpoint:** net adverse clinical and cerebral events
ie NACCE=death, MI, stroke or major bleed at 1 year

Expected 9%, non-inferiority margin 2.7%

<table>
<thead>
<tr>
<th>treatment duration</th>
<th>3 months [N=1563]</th>
<th>12 months [N=1556]</th>
<th>risk difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NACCE: at 1 year</td>
<td>93</td>
<td>90</td>
<td>0.2%(-1.5% to 1.9%)</td>
</tr>
<tr>
<td>: 3 m to 1 yr</td>
<td>54</td>
<td>52</td>
<td>0.1%(-1.25 to 1.4%)</td>
</tr>
<tr>
<td>stent thrombosis</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>major bleeding</td>
<td>3</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

trade-off between small risk of stent thrombosis
and small benefit of less bleeding

trial too small to answer that question

---

Non randomised studies: caution needed

drug eluting vs bare-metal stents for angioplasty

**Meta-analysis of RCTs and Registries**
[Circulation 2009;119:3198-3206]

**Hazard Ratios (DES vs BMS) for Mortality**

**21 RCTs** (8867 patients, mean f/u 2.9 years)
Fixed Effect 0.97 (95% CI 0.81, 1.15) P=.72

**31 Registries** (169,595 patients, mean f/u 2.5 years)
Fixed Effect 0.81 (95% CI 0.78, 0.85) P<.001
Random Effects 0.78 (95% CI 0.71, 0.86)

**RCTs:** no mortality difference
**Registries:** lower mortality after DES
Why such discrepancies between RCTs and Registries?

RCTs not representative of real-world use

Registries prone to selection bias,
   not captured by adjustment for confounders,
   which vary enormously across registries

mortality risk depends on so many factors
not related to specific PCI

any true effect (DES vs BMS) should be small?

Comparative Effectiveness of Revascularisation Strategies
[Weintraub et al, NEJM 2012;366:1467-76]

ASCERT CMS Registry in 2004 to 2008

103,549 PCI vs 86,244 CABG patients

age 65+, two-or three vessel disease without acute MI

median 2.67 years follow-up

~24,000 patients died
Propensity Score for CABG

Adjusted Rates of Survival using Inverse Probability Weighting

16.4% vs 20.8% dead at 4 years, RR 0.79 (95% CI 0.76-0.82)
Comparative Effectiveness of Multivessel PCI vs CABG  
[Hlatky et al Ann Intern Med 2013;158:727-34]

Medicare beneficiaries age 66+, median follow-up 4.3 years

105,156 propensity score-matched PCI/CABG patients

strongest predictors of receiving CABG rather than PCI:
calendar year, age, sex, diabetes, CVD, PAD, atrial fib, region

all cause mortality at 5 years 28.1% PCI, 25.9% CABG

hazard ratio 0.92 (95% CI 0.90 to 0.95) P<.001

RCTs of PCI vs CABG of limited size

**FREEDOM trial** in 1900 diabetics  
[NEJM 2012; 367 p 2375]

16.3% vs 10.9% mortality at 5 years  P=.049

**SYNTAX trial** in 1800 patients  
[Lancet 2013; 381 p 629-]

13.9% vs 11.4% mortality at 5 years  P=.10

We await the EXCEL trial in 2600 patients

meantime can we trust the registries?
A RESEARCH SURVIVAL KIT

Life as an academic medical statistician and how to survive it

[Stats in Medicine 1995; 14 p 209-222]

DO BOTH

Methods and Applications

Trials and Epidemiology
VARIETY

vs

EXPERT

strike a balance

MIX IN

Multi-disciplinary collaboration is fun
DOOR OPEN

opportunism

BIG PICTURE

don’t get lost in details
DEADLINES

make them…… again and again

ENJOY IT

back winners, drop losers

the art of saying NO
START

each day with the good stuff

LISTS

prioritize continuously
ENTHUSIASTS

work with them

teach them

share their money!

VISION

DRIVE

INSPIRATION