How clinical trial design pitfalls slow progress against cancer

David J. Stewart, MD, FRCPC

The Ottawa Hospital & University of Ottawa
Head, Division of Medical Oncology
Consulting/Advisory Boards:
- Roche Canada 2014
- Pfizer Canada 2014
- Boehringer Ingelheim Canada 2015
- Amgen/Amgen Canada 2014
- Novartis Canada 2015

Speaker:
- Pfizer Canada 2015
- IASLC 2014, 2015

Scientific writing support (review on angiogenesis): Boehringer Ingelheim 2015

Clinical trials support:
- Boehringer Ingelheim
- AstraZeneca
- Novartis
- Bristol-Myers Squibb
- Celgene
Question #1:
Why have gains against cancer been so small?
The late 60’s and 70’s were a time of great promise & Childhood ALL was “The paradigm for all cancers”
My prediction 1976

Within a few years, rapid advances in common cancers would mirror those seen in hematological and germ cell tumors.
The 2016 Reality

Most metastatic malignancies remain incurable.

“Statistical victories” with $p < 0.05$ are hailed as advances despite survival gains of mere weeks.
Recent “significant advances” (p<0.05) vs cancer

<table>
<thead>
<tr>
<th>Drug</th>
<th>Tumor</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>gemcitabine</td>
<td>pancreas</td>
<td>6 weeks</td>
</tr>
<tr>
<td>bevacizumab</td>
<td>colon</td>
<td>2.2 months</td>
</tr>
<tr>
<td>erlotinib</td>
<td>pancreas</td>
<td>11 days</td>
</tr>
<tr>
<td>bevacizumab</td>
<td>NSCLC</td>
<td>2 months</td>
</tr>
<tr>
<td>sorafenib</td>
<td>renal</td>
<td>2 months</td>
</tr>
<tr>
<td>temozolamide</td>
<td>GBM</td>
<td>2.5 months</td>
</tr>
<tr>
<td>docetaxel</td>
<td>prostate</td>
<td>2.4 months</td>
</tr>
<tr>
<td>cetuximab</td>
<td>colon</td>
<td>1.5 months</td>
</tr>
</tbody>
</table>
Problem even worse in cardiology: large trials are “significant” despite trivial gains

<table>
<thead>
<tr>
<th>Group</th>
<th>Rx</th>
<th>Endpoint</th>
<th>No.</th>
<th>%</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post MI</td>
<td>Rivaroxaban vs placebo</td>
<td>Cardiac death/MI/stroke</td>
<td>15,526</td>
<td>8.9 vs 10.7</td>
<td>0.008</td>
</tr>
<tr>
<td>CAD</td>
<td>Everolimus vs paclitaxel stent</td>
<td>Cardiac death/MI/ revascularization</td>
<td>3,687</td>
<td>4.2 vs 6.8</td>
<td>0.001</td>
</tr>
<tr>
<td>CAD</td>
<td>Zotarolimus vs sirolimus stent</td>
<td>Cardiac death/MI/ revascularization</td>
<td>2,332</td>
<td>6 vs 3</td>
<td>0.0002</td>
</tr>
<tr>
<td>Post MI</td>
<td>Ticagrelor vs clopidroegrel</td>
<td>Cardiac death/MI/stroke</td>
<td>18,624</td>
<td>9.8 vs 11.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Post MI</td>
<td>Pasugrel vs clopidroegrel</td>
<td>Cardiac death/MI/stroke</td>
<td>3,534</td>
<td>6.5 vs 9.5</td>
<td>0.0017</td>
</tr>
</tbody>
</table>

* “14% reduction in mortality”
Riveroxaban vs placebo in patients with a recent acute coronary syndrome

Mega, NEJM 366:9, 2012
Riveroxaban vs placebo in patients with a recent acute coronary syndrome

Mega, NEJM 366:9, 2012

n = 15,526

Hazard ratio, 0.84 (95% CI, 0.74–0.96)
P=0.008
Riveroxaban vs placebo in patients with a recent acute coronary syndrome

Hazard ratio, 0.84 (95% CI, 0.74–0.96)
P=0.008

n = 15,526
Why have we gained so little?

“Every system is perfectly designed to get exactly the results it gets!”

-P Batalden, F Davidoff, 2007

We have only gained little since RCTs are often specifically designed to detect small gains
Almost everyone except the patient gains from a low efficacy bar!

- Clinical investigator
- Statisticians
- Drug companies
- Granting agencies
- Regulators
- Health Care Providers
- Insurers
- Journals (Advertising)
- Politicians (Campaign $)
Common cancers: We have ensnared ourselves in a war of the trenches, aiming to advance yards, not miles!
General Oskar von Hutier and the 1918 Spring Offensive: small units, high goals, high mobility: stormed right past the trenches!
The Battle of Amiens, Aug 8, 1918: the start of the Hundred Days Offensive that led to Allied victory 3 months later.

“The importance of this offensive in tactical terms was that it .... provided a further example of the effectiveness of shifting from stagnant trench warfare to mobile multi-faceted warfare.”
• We need small studies aiming for big gains, not big studies aiming for small gains!
Question # 2: Are randomized controlled trials in unselected patients misleading?
Common cancers are common since they can be caused by many different mutations.

Hence: a targeted agent will work only in a small proportion of patients with a given common tumor type.

- Braiteh & Kurzrock, 2007
Targeted agents work in small subpopulations driven by specific mutations.
Shoeland: the tale of evidence-based shoe sales
Simulations illustrating how randomized trials in unselected patients may be misleading, unreliable and wasteful

- Used actual survival of 334 lung cancer patients as “control”
- Simulated different degrees of benefit for 334-patient “experimental group”
Therapy would be inappropriately abandoned if it hit a target present in only every 10th patient and quintupled their survival.

HR = 0.85, P = 0.16

Therapy would be inappropriately abandoned if it hit a target present in only every 10th patient and quintupled their survival.

668 patients “accrued”

HR = 0.85, P = 0.16

“Throw out drug since it is ineffective”

Study costs in unselected patients

Estimated cost/patient on phase III trials = $74,800
Cost for 668 patient “unselected” study: $50 M

- we squander patient research candidates
- we expose 90% patients to Rx that can’t help them
- and we get the wrong answer!

- “Drug is ineffective”
Tripling patient numbers to 2000 → magic!

↑ statistical power → $p < 0.03$

- New standard of care
- PI: ASCO plenary / NEJM / promotion / speakers circuit
- Study costs $150 M
- Still have wrong answer!
  - “Drug is effective.”
  - Widely used, including in the 90% of patients it cannot help

Large randomized trials in unselected patients are extraordinarily wasteful!
Markedly reduce required patient numbers and costs if select for patients with target

668 unselected patients, every 10\textsuperscript{th} has target:

HR = 0.85, P = 0.16

16 patients selected for target: p < 0.02

Cost to identify and study 10% (16) selected patients:

- biopsy/profiling ($\sim$10,000/pt x 160 pts) = $1.6 M
- study: 16 x $74,800 = $1.2 M
- Total: $2.8 M

-and you get data/other studies for other 90% pts

-and you may get right answer!
With target in 10% of patients, defining target & selecting patients molecularly:

- cuts no. patients entered by 98%
- cuts total costs (including screening) by 95%
Therapies that give very small gains in a high proportion of patients may be easier to detect than therapies giving marked benefit in subpopulations
Increasing survival of all patients by 33% achieved significance (HR = 0.80, p = 0.03)
Randomized trials may incorrectly conclude that 2 therapies are equivalent even though they hit completely different targets!
Therapies that work in completely different subpopulations would be declared to be "equivalent" (p=0.89)
Do not compare drugs hitting different targets!
Do not compare drugs hitting different targets!

- A is not “best”: it just hits a more common target!
Do not compare drugs hitting different targets!

- A is not “best”: it just hits a more common target!
- Since there is no target more common, no “better” drug will ever be found
Do not compare drugs hitting different targets!

- A is not “best”: it just hits a more common target!
- Since there is no target more common, no “better” drug will ever be found
- No progress is made since conclusions are wrong

Do not compare drugs hitting different targets!

- A is not “best”: it just hits a more common target!
- Since there is no target more common, no “better” drug will ever be found
- No progress is made since conclusions are wrong
- **Irrational** to compare agents hitting different targets

Very small change in % of patients with target may determine significance vs non-significance

Target in 15%:
HR = 0.81
p = 0.06

Target in 16.7%:
HR = 0.79
p = 0.04
Randomized trials in unselected patients may not detect that an agent is beneficial in one subpopulation but harmful in another.
Erlotinib didn’t improve PFS when added to NSCLC chemotherapy (TRIBUTE)

Median TTP (months)
Hazard ratio

<table>
<thead>
<tr>
<th></th>
<th>Erlotinib</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median TTP</td>
<td>5.1</td>
<td>4.9</td>
</tr>
<tr>
<td>Hazard ratio</td>
<td>0.937</td>
<td>0.940</td>
</tr>
<tr>
<td>P-value</td>
<td>0.36</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Herbst et al, 2005
TRIBUTE: Large erlotinib beneficial impact in EGFR mutants (13% of population, p=0.09) hidden by smaller detrimental impact in larger population of K-ras mutants (21% of population, p=0.03)
Conclusions

• Randomized trials in unselected patients:
  • Use large patient numbers to ↑ statistical power → detect small gains → set efficacy bar too low!
Conclusions

• Randomized trials in unselected patients:
  • Use large patient numbers to ↑ statistical power → detect *small* gains → set efficacy bar *too low*!
  • Discard therapies valuable in subpopulations
Conclusions

- Randomized trials in unselected patients:
  - Use large patient numbers to ↑ statistical power → detect small gains → set efficacy bar too low!
  - Discard therapies valuable in subpopulations
  - Foster broad use of therapies in subpopulations that can not benefit
Conclusions

• Randomized trials in unselected patients:
  • Use large patient numbers to ↑ statistical power → detect *small* gains → set efficacy bar *too low!*
  • Discard therapies valuable in subpopulations
  • Foster broad use of therapies in subpopulations that can not benefit

• Inordinately expensive
Change the question
Change the question: “Who benefits?!”

Not: “Is there an average benefit?”
Question # 3
Can we afford routine broad molecular profiling?
Question # 3
Can we afford routine broad molecular profiling?

Glieberman principle:

“Hire the best coach that money can buy. The second best is too expensive!”

While costs of molecular assessments may be high, long term costs of “traditional” approaches are much higher!
Average cost of 4 weeks of Rx in Canada

Anticancer agents = 31% of patented drug sales in Canada
Question # 4:

Should overall survival remain as our gold standard primary endpoint in cancer clinical trials?
Advantages of survival as endpoint

• Measured precisely
• Important to patients
• Current “gold standard”
But overall survival has less statistical power than progression-free survival or response.
Hazard ratio ~ 0.5
Overall Survival with PFS 6 months vs 3 months, but Survival Post Progression 18 Months Both Arms

Hazard ratio ~ 0.88
Cross-over can have a major negative impact on conclusions of studies using overall survival (rather than PFS or response) as the primary endpoint.
Impact of cross over

No crossover (2,000 patients):
P = 0.0148
HR = 0.85
Control median OS = 77.9

Crossover of 20% of patients with target (2% of total population):
P = 0.069
HR = 0.88
Control median OS = 78.7
Impact of cross over may be even higher if patients were selected based on predictive biomarker

\[(n=16, p<0.02, HR = 0.16)\]

\[(n=16, p = 0.07, HR = 0.28)\]
Vemurafenib much more effective than chemotherapy in giving tumor regression / symptom improvement in BRAF-mutant melanoma patients

Chemotherapy

Vemurafenib:
>95% had regression
Impact of Vemurafenib on progression-free and overall survival in advanced melanoma with BRAF V600E mutation

P Chapman et al NEJM 2011
Growing up in California’s rural Central Valley, the two cousins spent summers racing dirt bikes and Christmases at their grandmother’s on the coast. Endowed with a similar brash charm, they bought each other matching hardhats and sought iron-working jobs together. They shared a love for the rush that comes with hanging steel at dizzying heights, and a knack for collecting speeding tickets.

And when, last year, each learned that a lethal skin cancer called melanoma was spreading rapidly through his body, the young men found themselves with the shared chance of benefiting from a recent medical breakthrough.
Impact of cross-over on overall survival

• Overall survival (but not PFS or response) affected by cross-over:
  • To the investigational therapy
  • *To any other effective therapy*
  • To Palliative care
Impact of cross-over on overall survival

• Overall survival (but not PFS or response) affected by cross-over:
  • To the investigational therapy
  • To any other effective therapy
  • To Palliative care
→ False negative study
Impact of cross-over on overall survival

- Overall survival (but not PFS or response) affected by cross-over:
  - To the investigational therapy
  - *To any other effective therapy*
  - To Palliative care

→ False negative study

→ Temptation for unethical trial design (restricted cross over)
Question # 5:  

Can response be a valid primary endpoint in cancer clinical trials?
False positives are uncommon in single agent studies

- % of patients with RECIST Partial Response on placebo or best supportive care arms of randomized trials:
  - median 1% (range, 0-4%)
Single agent response predicts effective drug

- Response rate in single agent phase II predicts high probability of regulatory approval ($p<0.005$)$^1$
- Response correlates with survival across trials, $p<0.0001$$^2,3$

References
Tumor regression in > 90% of patients with target

Vemurafenib: BRAF-mutant melanoma

Erlotinib/gefitinib in EGFR mutant NSCLC

Crizotinib in EML4/ALK fusion NSCLC

Crizotinib in ROS1 fusion NSCLC
Of 31 drugs approved by FDA based on response (non-randomized) since 1970, 30 remain approved.
To improve precision of response measurement

- Measure larger tumors
- Increase number of tumors measured, if small
- Volumetric assessment
- Side-by-side measurements of pre and post scans
Conclusion: reasonable to approve a drug based on a high single-agent response rate in disease that is refractory to standard therapy
Question #6:

Can Progression-Free Survival (PFS) be a valid primary endpoint in cancer clinical trials?
Overall Survival vs PFS vs Response as Primary Study Endpoint

<table>
<thead>
<tr>
<th></th>
<th>Survival</th>
<th>Progression-Free Survival</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Best</td>
<td>Least</td>
<td>Less</td>
</tr>
<tr>
<td>Impact of post-progression survival</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Impact of cross-over</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Statistical power</td>
<td>Lowest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
<tr>
<td>Randomized trial needed</td>
<td>Yes</td>
<td>Yes</td>
<td>Not if single agent</td>
</tr>
<tr>
<td>No. patients needed*</td>
<td>Most</td>
<td>Fewer</td>
<td>Least</td>
</tr>
<tr>
<td>Follow-up time*</td>
<td>Longest</td>
<td>Shorter</td>
<td>Shortest</td>
</tr>
<tr>
<td>Total costs*</td>
<td>Most</td>
<td>Less</td>
<td>Least</td>
</tr>
<tr>
<td>Biomarker discovery</td>
<td>Poorest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
</tbody>
</table>

* Translates into delay in drug approval & higher drug costs → ↑ life-years lost, ↑ patient suffering, ↑ drug prices / ↑ health care costs
Overall Survival vs PFS vs Response as Primary Study Endpoint

<table>
<thead>
<tr>
<th></th>
<th>Survival</th>
<th>Progression-Free Survival</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Best</td>
<td>Least</td>
<td>Less</td>
</tr>
<tr>
<td>Impact of post-progression survival</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Impact of cross-over</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Statistical power</td>
<td>Lowest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
<tr>
<td>Randomized trial needed</td>
<td>Yes</td>
<td>Yes</td>
<td>Not if single agent</td>
</tr>
<tr>
<td>No. patients needed*</td>
<td>Most</td>
<td>Fewer</td>
<td>Least</td>
</tr>
<tr>
<td>Follow-up time*</td>
<td>Longest</td>
<td>Shorter</td>
<td>Shortest</td>
</tr>
<tr>
<td>Total costs*</td>
<td>Most</td>
<td>Less</td>
<td>Least</td>
</tr>
<tr>
<td>Biomarker discovery</td>
<td>Poorest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
</tbody>
</table>

* Translates into delay in drug approval & higher drug costs → ↑ life-years lost, ↑ patient suffering, ↑ drug prices / ↑ health care costs
### Overall Survival vs PFS vs Response as Primary Study Endpoint

<table>
<thead>
<tr>
<th></th>
<th>Survival</th>
<th>Progression-Free Survival</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Best</td>
<td>Least</td>
<td>Less</td>
</tr>
<tr>
<td>Impact of post-progression survival</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Impact of cross-over</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Statistical power</td>
<td>Lowest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
<tr>
<td>Randomized trial needed</td>
<td>Yes</td>
<td>Yes</td>
<td>Not if single agent</td>
</tr>
<tr>
<td>No. patients needed*</td>
<td>Most</td>
<td>Fewer</td>
<td>Least</td>
</tr>
<tr>
<td>Follow-up time*</td>
<td>Longest</td>
<td>Shorter</td>
<td>Shortest</td>
</tr>
<tr>
<td>Total costs*</td>
<td>Most</td>
<td>Less</td>
<td>Least</td>
</tr>
<tr>
<td>Biomarker discovery</td>
<td>Poorest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
</tbody>
</table>

* Translates into delay in drug approval & higher drug costs → ↑ life-years lost, ↑ patient suffering, ↑ drug prices / ↑ health care costs
Overall Survival vs PFS vs Response as Primary Study Endpoint

<table>
<thead>
<tr>
<th></th>
<th>Survival</th>
<th>Progression-Free Survival</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Best</td>
<td>Least</td>
<td>Less</td>
</tr>
<tr>
<td>Impact of post-progression survival</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Impact of cross-over</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Statistical power</td>
<td>Lowest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
<tr>
<td>Randomized trial needed</td>
<td>Yes</td>
<td>Yes</td>
<td>Not if single agent</td>
</tr>
<tr>
<td>No. patients needed*</td>
<td>Most</td>
<td>Fewer</td>
<td>Least</td>
</tr>
<tr>
<td>Follow-up time*</td>
<td>Longest</td>
<td>Shorter</td>
<td>Shortest</td>
</tr>
<tr>
<td>Total costs*</td>
<td>Most</td>
<td>Less</td>
<td>Least</td>
</tr>
<tr>
<td>Biomarker discovery</td>
<td>Poorest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
</tbody>
</table>

* Translates into delay in drug approval & higher drug costs → ↑ life-years lost, ↑ patient suffering, ↑ drug prices / ↑ health care costs
**Overall Survival vs PFS vs Response as Primary Study Endpoint**

<table>
<thead>
<tr>
<th></th>
<th>Survival</th>
<th>Progression-Free Survival</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Best</td>
<td>Least</td>
<td>Less</td>
</tr>
<tr>
<td>Impact of post-progression survival</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Impact of cross-over</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Statistical power</td>
<td>Lowest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
<tr>
<td>Randomized trial needed</td>
<td>Yes</td>
<td>Yes</td>
<td>Not if single agent</td>
</tr>
<tr>
<td>No. patients needed*</td>
<td>Most</td>
<td>Fewer</td>
<td>Least</td>
</tr>
<tr>
<td>Follow-up time*</td>
<td>Longest</td>
<td>Shorter</td>
<td>Shortest</td>
</tr>
<tr>
<td>Total costs*</td>
<td>Most</td>
<td>Less</td>
<td>Least</td>
</tr>
<tr>
<td>Biomarker discovery</td>
<td>Poorest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
</tbody>
</table>

* Translates into delay in drug approval & higher drug costs → ↑ life-years lost, ↑ patient suffering, ↑ drug prices / ↑ health care costs
Overall Survival vs PFS vs Response as Primary Study Endpoint

<table>
<thead>
<tr>
<th></th>
<th>Survival</th>
<th>Progression-Free Survival</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Best</td>
<td>Least</td>
<td>Less</td>
</tr>
<tr>
<td>Impact of post-progression survival</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Impact of cross-over</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Statistical power</td>
<td>Lowest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
<tr>
<td>Randomized trial needed</td>
<td>Yes</td>
<td>Yes</td>
<td>Not if single agent</td>
</tr>
<tr>
<td>No. patients needed*</td>
<td>Most</td>
<td>Fewer</td>
<td>Least</td>
</tr>
<tr>
<td>Follow-up time*</td>
<td>Longest</td>
<td>Shorter</td>
<td>Shortest</td>
</tr>
<tr>
<td>Total costs*</td>
<td>Most</td>
<td>Less</td>
<td>Least</td>
</tr>
<tr>
<td>Biomarker discovery</td>
<td>Poorest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
</tbody>
</table>

* Translates into delay in drug approval & higher drug costs → ↑ life-years lost, ↑ patient suffering, ↑ drug prices / ↑ health care costs
# Overall Survival vs PFS vs Response as Primary Study Endpoint

<table>
<thead>
<tr>
<th></th>
<th>Survival</th>
<th>Progression-Free Survival</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Best</td>
<td>Least</td>
<td>Less</td>
</tr>
<tr>
<td>Impact of post-progression survival</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Impact of cross-over</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Statistical power</td>
<td>Lowest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
<tr>
<td>Randomized trial needed</td>
<td>Yes</td>
<td>Yes</td>
<td>Not if single agent</td>
</tr>
<tr>
<td>No. patients needed*</td>
<td>Most</td>
<td>Fewer</td>
<td>Least</td>
</tr>
<tr>
<td>Follow-up time*</td>
<td>Longest</td>
<td>Shorter</td>
<td>Shortest</td>
</tr>
<tr>
<td>Total costs*</td>
<td>Most</td>
<td>Less</td>
<td>Least</td>
</tr>
<tr>
<td>Biomarker discovery</td>
<td>Poorest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
</tbody>
</table>

* Translates into delay in drug approval & higher drug costs → ↑ life-years lost, ↑ patient suffering, ↑ drug prices / ↑ health care costs
# Overall Survival vs PFS vs Response as Primary Study Endpoint

<table>
<thead>
<tr>
<th></th>
<th>Survival</th>
<th>Progression-Free Survival</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Best</td>
<td>Least</td>
<td>Less</td>
</tr>
<tr>
<td>Impact of post-progression survival</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Impact of cross-over</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Statistical power</td>
<td>Lowest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
<tr>
<td>Randomized trial needed</td>
<td>Yes</td>
<td>Yes</td>
<td>Not if single agent</td>
</tr>
<tr>
<td>No. patients needed*</td>
<td>Most</td>
<td>Fewer</td>
<td>Least</td>
</tr>
<tr>
<td>Follow-up time*</td>
<td>Longest</td>
<td>Shorter</td>
<td>Shortest</td>
</tr>
<tr>
<td>Total costs*</td>
<td>Most</td>
<td>Less</td>
<td>Least</td>
</tr>
<tr>
<td>Biomarker discovery</td>
<td>Poorest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
</tbody>
</table>

* Translates into delay in drug approval & higher drug costs → ↑ life-years lost, ↑ patient suffering, ↑ drug prices / ↑ health care costs
### Overall Survival vs PFS vs Response as Primary Study Endpoint

<table>
<thead>
<tr>
<th></th>
<th>Survival</th>
<th>Progression-Free Survival</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Best</td>
<td>Least</td>
<td>Less</td>
</tr>
<tr>
<td>Impact of post-progression survival</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Impact of cross-over</td>
<td>Most</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Statistical power</td>
<td>Lowest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
<tr>
<td>Randomized trial needed</td>
<td>Yes</td>
<td>Yes</td>
<td>Not if single agent</td>
</tr>
<tr>
<td>No. patients needed*</td>
<td>Most</td>
<td>Fewer</td>
<td>Least</td>
</tr>
<tr>
<td>Follow-up time*</td>
<td>Longest</td>
<td>Shorter</td>
<td>Shortest</td>
</tr>
<tr>
<td>Total costs*</td>
<td>Most</td>
<td>Less</td>
<td>Least</td>
</tr>
<tr>
<td>Biomarker discovery</td>
<td>Poorest</td>
<td>Intermediate</td>
<td>Best</td>
</tr>
</tbody>
</table>

* Translates into delay in drug approval & higher drug costs → ↑ life-years lost, ↑ patient suffering, ↑ drug prices / ↑ health care costs
Question #7:

With Progression-Free Survival as primary endpoint, how big a gain should you aim for?
Gain in PFS t1/2 vs Overall Survival t1/2 with Bevacizumab compared to controls: 28 studies, 9 tumor types*

* Excludes studies with cross-over and 2 studies with outliers / immature OS data
Gain in PFS t1/2 vs Overall Survival t1/2) with Bevacizumab compared to controls: 28 studies, 9 tumor types*

* Excludes studies with cross-over and 2 studies with outliers / immature OS data

D Bosse: unpublished
Gain in PFS t1/2 vs Overall Survival t1/2) with Bevacizumab compared to controls: 28 studies, 9 tumor types*

1. Gain in PFS t1/2 (in months, compared to control) was similar to the gain in overall survival t1/2 (median, 1.7 months for both).

D Bosse: unpublished

* Excludes studies with cross-over and 2 studies with outliers / immature OS data
1. Gain in PFS t1/2 (in months, compared to control) was similar to the gain in overall survival t1/2 (median, 1.7 months for both).

2. Does this also hold for other drugs?

D Bosse: unpublished

* Excludes studies with cross-over and 2 studies with outliers / immature OS data
Gain in PFS t1/2 vs Overall Survival t1/2) with Bevacizumab compared to controls: 28 studies, 9 tumor types*

1. Gain in PFS t1/2 (in months, compared to control) was similar to the gain in overall survival t1/2 (median, 1.7 months for both).

2. Does this also hold for other drugs?

3. Less variability in PFS vs OS (impact of post-progression survival)

* Excludes studies with cross-over and 2 studies with outliers / immature OS data
Possible approach for use of PFS as primary endpoint for drug approval

• Reach consensus on what would be desired absolute gain (months) in overall survival
Possible approach for use of PFS as primary endpoint for drug approval

- Reach consensus on what would be desired absolute gain (months) in overall survival

- Definition of positive study:
  - Lower limit of 95% confidence intervals for observed PFS gain = desired overall survival gain
Target Survival gain is 4 months

Target survival gain = 4 months
Target Survival gain is 4 months: PFS 95% Confidence Intervals are calculated

Target survival gain = 4 months

PFS gain
95% CI
Drug is discarded if lower boundary of PFS 95% Confidence Intervals is less than 4 months

Target survival gain = 4 months
Target Survival gain is 4 months: PFS 95% Confidence Intervals are calculated

Target survival gain = 4 months
Drug is approved if lower boundary of PFS 95% Confidence Intervals is greater than 4 months

Target survival gain = 4 months
Question #8:

When is a randomized trial essential?
When is randomized trial needed?

• Randomized trial not needed:
  • High response rate with single agent in refractory population
When is randomized trial needed?

• Randomized trial **not** needed:
  • High response rate with single agent in refractory population

• Randomization generally **needed**:
  • Overall survival or Progression-free survival endpoint
  • Drug combinations
  • High anticipated response rate with standard therapy?
Question #9:

Why are usual designs for studies of drug combinations problematic?
Combinations: problematic

- Trials may be unreliable & misleading, whether:
  - Non-randomized
  - Randomized (using current trial methods)
If you do not randomize, you cannot tell whether apparent benefit from B is real (vs variability in efficacy of A).
But randomized trials can also mislead re value of combinations

- Combination more effective than either single agent, but:
  - all patients get at least 1 drug that doesn’t help them
  - 55% get 2 drugs that don’t help

Combination vs sequential single agent chemotherapy in breast cancer

<table>
<thead>
<tr>
<th>Chemo</th>
<th>Response rate</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combination</td>
<td>46%</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Sequential</td>
<td>25%</td>
<td></td>
</tr>
</tbody>
</table>

Chlebowski et al, BJC 1989
How studies of combinations can mislead if patients cross over to the new drug at progression

- A+B may have:
  - Significant ↑ response rate & time to progression
  - No ↑ overall survival

How studies of combinations can mislead if patients cross over to the new drug at progression

- $A+B$ may have:
  - Significant $\uparrow$ response rate & time to progression
  - No $\uparrow$ overall survival
  - $B$ may be discarded as “ineffective” since it is as effective 2nd line as 1st line

How studies of combinations can mislead if patients cross over to the new drug at progression

- A+B may have:
  - Significant ↑ response rate & time to progression
  - No ↑ overall survival
- B may be discarded as “ineffective” since it is as effective 2\textsuperscript{nd} line as 1\textsuperscript{st} line
- Study quoted as example of response rate & time to progression not being reliable surrogates of overall survival

How studies of combinations can mislead if patients cross over to the new drug at progression

- Never discard B based solely on a negative combination trial: potential impact of →
  - Antagonism with A
  - Cross-over
  - Other factors that ↑ post-progression survival

How studies of combinations can mislead if *no* cross over to the new drug at progression

- A+B may have:
  - Significant ↑ response rate & time to progression
  - Significant ↑ overall survival

How studies of combinations can mislead if no cross over to the new drug at progression

A+B may have:

- Significant ↑ response rate & time to progression
- Significant ↑ overall survival

A+B becomes new “standard of care”
How studies of combinations can mislead if no cross over to the new drug at progression

- A+B may have:
  - Significant ↑ response rate & time to progression
  - Significant ↑ overall survival
- A+B becomes new “standard of care”
- Study quoted as example of response rate & time to progression being reliable surrogates of overall survival

How studies of combinations can mislead if _no_ cross over to the new drug at progression

- A+B may have:
  - Significant ↑ response rate & time to progression
  - Significant ↑ overall survival
- A+B becomes new “standard of care”
- Study quoted as example of response rate & time to progression being reliable surrogates of overall survival
- Problem: which patients need which drug(s)?

How studies of combinations can mislead if no cross over to the new drug at progression

- A+B may have:
  - Significant ↑ response rate & time to progression
  - Significant ↑ overall survival
- A+B becomes new “standard of care”
- Study quoted as example of response rate & time to progression being reliable surrogates of overall survival
- Problem: which patients need which drug(s)?
- Unethical to block cross-over?

Combinations: Proposed design to assess if A+B is **really** better than A or B alone or sequentially?

- Maximum regression/response
- Total time to final progression
- Overall survival

Combinations: Proposed design to assess if A+B is really better than A or B alone or sequentially?

If there is true synergism then A+B should be much better than sequential

Question #10:

What is the best method to discover predictive biomarkers?
Response is better than survival as endpoint for discovery of predictive biomarkers

• ↓ time required (weeks follow-up, not months/years)
Response is better than survival as endpoint for discovery of predictive biomarkers

- ↓ time required (weeks follow-up, not months/years)
- ↓ patients needed → ↑ no. biomarkers that can be assessed:
  - Small single arm study: randomization not needed to differentiate predictive from prognostic markers
Response is better than survival as endpoint for discovery of predictive biomarkers

- ↓ time required (weeks follow-up, not months/years)
- ↓ patients needed → ↑ no. biomarkers that can be assessed:
  - Small single arm study: randomization not needed to differentiate predictive from prognostic markers
  - ↑ statistical power: correlation of marker with outcome is stronger for response / weaker for survival since response not diluted by:
    - Post-progression survival
    - Cross over to study drug

Higher statistical power: stronger correlation of biomarker with response vs survival

<table>
<thead>
<tr>
<th>Tumor</th>
<th>Therapy</th>
<th>Biomarker</th>
<th>No. studies: stronger correlation with response</th>
<th>No. studies: stronger correlation with survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon</td>
<td>cetuximab/panitumumab</td>
<td>KRAS mutation</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>NSCLC</td>
<td>gefitinib/erlotinib</td>
<td>EGFR mutation</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>
Panitumumab in colon Ca: needed 391 patients to show KRAS mutation impact using survival outcome vs 24 with response outcome

\[ \text{KRAS mutant (184 patients)} \]

\[ \text{KRAS wild-type (207 patients)} \]

<table>
<thead>
<tr>
<th>Shrinkage KRAS wild type</th>
<th>55% of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shrinkage KRAS mutant</td>
<td>4% of patients</td>
</tr>
<tr>
<td>No. patients needed for ( p &lt; 0.05 )</td>
<td>(~12) of each</td>
</tr>
</tbody>
</table>

R. Amado et al JCO 2008

OHRI IRHO
EGFR mutation detected as a target for gefitinib by comparing 5 responders to 4 non-responders from phase II trials.

<table>
<thead>
<tr>
<th>EGFR mutation</th>
<th>Responder</th>
<th>Non-responder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>No</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

p<0.0027

Paez et al 2004
# Number of patients required to identify target: response endpoint

<table>
<thead>
<tr>
<th>Tumor Type</th>
<th>Drug</th>
<th>Target</th>
<th>No. patients required</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSCLC</td>
<td>Gefitinib</td>
<td>$EGFR$</td>
<td>9</td>
</tr>
<tr>
<td>NSCLC</td>
<td>Crizotinib</td>
<td>$EML4/ALK$</td>
<td>2</td>
</tr>
</tbody>
</table>
Tumor regression in > 90% of patients with target

Vemurafenib: BRAF-mutant melanoma

Erlotinib/gefitinib in EGFR mutant NSCLC

Crizotinib in EML4/ALK fusion NSCLC

Crizotinib in ROS1 fusion NSCLC
### Drug in marker-positive groups vs drug in marker-negative groups vs Placebo / BSC

<table>
<thead>
<tr>
<th>Group</th>
<th>Response rate</th>
<th>% with &gt;10% tumor regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marker positive</td>
<td>70% (25-84)</td>
<td>86% (50-95)</td>
</tr>
<tr>
<td>Marker negative</td>
<td>6% (0-10)</td>
<td>20% (1-38)</td>
</tr>
<tr>
<td>Placebo / BSC</td>
<td>1% (0-4)</td>
<td>6% (0-9)</td>
</tr>
</tbody>
</table>
Biomarker discovery

- Mutations / amplifications better than gene or protein expression:

Biomarker discovery

• Mutations / amplifications better than gene or protein expression:
  • Will not vary:
    • Over short time frame (eg, diurnal)
    • In response to external stimuli
Biomarker discovery

- Mutations / amplifications better than gene or protein expression:
  - Will not vary:
    - Over short time frame (e.g., diurnal)
    - In response to external stimuli
  - Dichotomous rather than continuous:
    - Few examples of useful continuous biomarkers unless “Almost none vs any”
Correlation of predictive biomarker "x" with survival: $r = 0.91$, $p < 0.0001$
Issues:

1. No matter where the cut point, group above will be better than group below
Issues:

1. No matter where the cut point, group above will be better than group below

2. Suggests that patient at 51st percentile is different than one at 49th percentile and same as one at 99th percentile: biologically irrational
Issues:

1. No matter where the cut point, group above will be better than group below

2. Suggests that patient at 51st percentile is different than one at 49th percentile and same as one at 99th percentile: biologically irrational

3. Only way it makes sense is to assign a probability of benefit rather than to divide patients into “good risk” vs “poor risk” groups
Question #11:

What is the problem when your top objective is to prove a point, not to help the patient?
New Drugs Stir Debate on Rules of Clinical Trials

By AMY HARMON SEPT. 18, 2010

Growing up in California’s rural Central Valley, the two cousins spent summers racing dirt bikes and Christmases at their grandmother’s on the coast. Endowed with a similar brash charm, they bought each other matching hardhats and sought ironworking jobs together. They shared a love for the rush that comes with hanging steel at dizzying heights, and a knack for collecting speeding tickets.

And when, last year, each learned that a lethal skin cancer called melanoma was spreading rapidly through his body, the young men found themselves with the shared chance of benefiting from a recent medical breakthrough.

P Chapman et al NEJM 2011

Cross-over forbidden
Denial of active therapy: Adjuvant chemo vs observation in children with osteogenic sarcoma
Don’t try to claim there is equipoise when there clearly is not.
Question #12:
Why are shorter, faster studies so very important?
Big, expensive studies drive high drug costs
Average cost of 4 weeks of Rx in Canada

Anticancer agents = 31% of patented drug sales in Canada
Big studies take longer / cost lives
Methods

• As illustrative examples we assessed 21 drugs yielding a survival advantage in 11 malignancies in 27 phase III trials published 2000-2015
Stewart et al. WCLC 2015

Median (range) improvement in median survival:
0.31 (0.12-1.31) years
Life-years lost worldwide per year delay in drug approval

Total combined life-years lost / year:

Worldwide: 2,541,274
North America: 240,085

1 for every 12 seconds delay

Stewart et al. WCLC 2015

Life-years lost worldwide per year
Years from US patent application to US FDA approval

Median (range) time drug discovery to approval:
12 (6.1-23.3) years

Stewart et al. WCLC 2015
Life-years lost worldwide from patent application to approval

Total combined life-years lost:
Worldwide: 31,537,958
North America: 2,956,667

Stewart et al. WCLC 2015
Conclusions

- Small studies aiming for large gains, not large studies aiming for small gains
Conclusions

• Small studies aiming for large gains, not large studies aiming for small gains

• Randomized trials in unselected patients are misleading
Conclusions

- Small studies aiming for large gains, not large studies aiming for small gains
- Randomized trials in unselected patients are misleading
- Don’t compare agents hitting different targets
Conclusions

• Small studies aiming for large gains, not large studies aiming for small gains
• Randomized trials in unselected patients are misleading
• Don’t compare agents hitting different targets
• We cannot afford to not do tumor molecular testing
Conclusions

- Small studies aiming for large gains, not large studies aiming for small gains
- Randomized trials in unselected patients are misleading
- Don’t compare agents hitting different targets
- We cannot afford to not do tumor molecular testing
- Progression-free survival and response have distinct advantages over overall survival as a trial endpoint: need defined decision rules
Conclusions

- Small studies aiming for large gains, not large studies aiming for small gains
- Randomized trials in unselected patients are misleading
- Don’t compare agents hitting different targets
- We cannot afford to not do tumor molecular testing
- Progression-free survival and response have distinct advantages over overall survival as a trial endpoint: need defined decision rules
- Randomized trials are not needed if there is a high single-agent response rate in refractory patients
Conclusions

• Small studies aiming for large gains, not large studies aiming for small gains
• Randomized trials in unselected patients are misleading
• Don’t compare agents hitting different targets
• We cannot afford to **not** do tumor molecular testing
• Progression-free survival and response have distinct advantages over overall survival as a trial endpoint: need defined decision rules
• Randomized trials **are not** needed if there is a high single agent response rate in refractory patients
• Randomized trials **are** needed for a survival or PFS endpoint or for drug combinations
Conclusions

• Small studies aiming for large gains, not large studies aiming for small gains
• Randomized trials in unselected patients are misleading
• Don’t compare agents hitting different targets
• We cannot afford to not do tumor molecular testing
• Progression-free survival and response have distinct advantages over overall survival as a trial endpoint: need defined decision rules
• Randomized trials are not needed if there is a high single agent response rate in refractory patients
• Randomized trials are needed for a survival or PFS endpoint or for drug combinations
• New trial design is needed to assess drug combinations
Conclusions

- Small studies aiming for large gains, not large studies aiming for small gains
- Randomized trials in unselected patients are misleading
- Don’t compare agents hitting different targets
- We cannot afford to not do tumor molecular testing
- Progression-free survival and response have distinct advantages over overall survival as a trial endpoint: need defined decision rules
- Randomized trials are not needed if there is a high single agent response rate in refractory patients
- Randomized trials are needed for a survival or PFS endpoint or for drug combinations
- New trial design is needed to assess drug combinations
- Best endpoint to discover biomarkers: single agent response
Conclusions

• Small studies aiming for large gains, not large studies aiming for small gains
• Randomized trials in unselected patients are misleading
• Don’t compare agents hitting different targets
• We cannot afford to not do tumor molecular testing
• Progression-free survival and response have distinct advantages over overall survival as a trial endpoint: need defined decision rules
• Randomized trials are not needed if there is a high single agent response rate in refractory patients
• Randomized trials are needed for a survival or PFS endpoint or for drug combinations
• New trial design is needed to assess drug combinations
• Best endpoint to discover biomarkers: single agent response
• Think carefully about the ethics of your study
Conclusions

• Small studies aiming for large gains, not large studies aiming for small gains
• Randomized trials in unselected patients are misleading
• Don’t compare agents hitting different targets
• We cannot afford to not do tumor molecular testing
• Progression-free survival and response have distinct advantages over overall survival as a trial endpoint: need defined decision rules
• Randomized trials are not needed if there is a high single agent response rate in refractory patients
• Randomized trials are needed for a survival or PFS endpoint or for drug combinations
• New trial design is needed to assess drug combinations
• Best endpoint to discover biomarkers: single agent response
• Think carefully about the ethics of your study
• Shorter, cheaper studies for rapid, inexpensive drug approval
References


5. The urgent need for clinical research reform to permit faster, less expensive access to new therapies for lethal diseases. Clin Cancer Res 21:4561, 2015

6. Impact of time to drug approval on potential years of life lost: the compelling need for improved trial and regulatory efficiency. 16th World Conf Lung Cancer abst # 3385, 2015