Issues with the design and interpretation of phase III clinical trials

Ian F. Tannock

Princess Margaret Hospital and University of Toronto

May-25-17
Potential conflicts of interest

I have advised multiple companies about design of clinical trials for which I have received contributions to my research fund.

I have chaired international company-sponsored trials for hormone-refractory prostate cancer (TAX-327, VENICE)

I do not speak for companies and do not participate in Pharma-sponsored satellite symposia

May-25-17
Which factors might limit relevance of RCTs?

RCTs prevent bias in comparing treatments and provide a sound basis for changes in clinical practice.

BUT, poorly-designed RCTs may:

1. Ask questions of commercial rather than clinical interest
2. Be based on inadequate preclinical and early clinical studies
3. Use surrogate endpoints that do not reflect patient benefit
4. Show statistically significant but clinically irrelevant results
5. Be analyzed and reported prematurely
6. Underestimate toxicity of new treatments
7. Be subject to biased reporting
8. Select patients who do not represent those seen in every-day practice.
1. RCTs may ask questions of commercial rather than clinical interest...

... and commercial interest means that treatment must be profitable, and not necessarily cost-effective.
• Number and size of RCTs increased with time (1975-2004)
• Increase of time to event measures (OS, PFS etc.) as the primary endpoint
• For profit sponsorship increased from 4% to 57%, and is now >80%
• Authors more likely to endorse the experimental arm, despite stable effect size.
• Significant p-value and industry sponsorship associated with endorsement of the experimental arm
There is no incentive to evaluate old (and cheaper) drugs for new indications

Why was premetrexed (€€€) evaluated for mesothelioma?

Answer: Because responses were seen with methotrexate CR+PR among 63 patients in a Norwegian trial: (Solheim et al. Br J Cancer 1992;65:956-60)

Is methotrexate as effective as premetrexed?

Is paclitaxel (€) as effective as nab-paclitaxel (€€€€) when used with gemcitabine for pancreatic cancer?

We will never know the answers, because there is no incentive to ask such questions in an RCT

May-25-17
It has become difficult to do trials that will not make a profit for a company.

Women with DCIS:
- ~2000 patients with DCIS of the breast
- Usual care (surgery or lumpectomy + RT)
- Observation

Women with colorectal cancer:
- ~2600 patients with colorectal cancer
- Aspirin 200mg/day for 3 years
- Matched placebo
- Usual surgery and adjuvant therapy

These trials, with huge potential benefit, MUST BE ALLOWED TO CONTINUE.
It is easier to perform a clinical trial...

...that compares similar strategies even if the results are of minimal interest...

...than to perform a trial of fundamental importance that compares quite different strategies

May-25-17
The obscene cost of new drugs: does it relate to their effectiveness?

Cost per Life Year gained:

If new treatment B costs €50,000/yr & standard treatment A costs €5,000/yr.

In an RCT patients live a median 6 months longer with B

Then added cost per Life-Year gained is €90,000

What cost per LY gained is cost-effective?

– In Toronto or Madrid, health-care systems can afford up to ~€100,000 per LY gained

– In developing countries health-care systems can afford much less

May-25-17
• We estimated the cost per life-year (LY) gained for 25 new drugs approved by FDA in 2000-2010
• For only 37% of new agents was the cost per LY gained <€100,000
• Cost of new drugs should be reduced by median 78% to be cost-effective, even for Western countries
• We suggest that registration of new anticancer drugs require value-based pricing that renders them cost-effective
Three groups of agents FDA approved since 2000:

(A) Targeted agents where target population is selected by a biomarker
(B) Less specific biological targeted agents
(C) Chemotherapy

<table>
<thead>
<tr>
<th>Group</th>
<th>No of drugs/trials</th>
<th>Median HR for OS</th>
<th>Median monthly cost (in USA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6/7</td>
<td>0.69</td>
<td>$5,400</td>
</tr>
<tr>
<td>B</td>
<td>7/14</td>
<td>0.78</td>
<td>$5,600</td>
</tr>
<tr>
<td>C</td>
<td>8/12</td>
<td>0.84</td>
<td>$6,600</td>
</tr>
</tbody>
</table>

P-value 0.003

There is no relationship between cost and value
Value-based pricing?

- New drugs should be priced within a limit of cost effectiveness (e.g. <€100,000 per life-year gained)
- Borderline drugs giving small differences in outcome would have to be cheap
- Effective drugs (e.g. trastuzumab, imatinib) can be expensive but still cost-effective.
If you are buying a....

Ford

You shouldn’t have to pay for a...

Ferrari
Questions (1)

.... about commercial drug development and cost effectiveness
Which factors might limit relevance of RCTs?

2. Poorly-designed RCTs may be based on inadequate preclinical and early clinical studies
There is substantial evidence that development of anti-cancer drugs is inefficient compared to those in other areas of medicine.

- 40 drugs were dropped from the global oncology pipeline in 2013.
- 20 drug terminations occurred in Phase 1 with reasons for termination commonly reported as strategic or undisclosed.
- 12 drugs (30%) failed in Phase 3.
- No pivotal trials for these agents incorporated molecular biomarkers for stratification.
Pharmacodynamics (PD)

**PD:** Is the drug doing what you think it is – is it inhibiting its putative molecular target?

- Important to study PD:
  - In preclinical studies (universally done but problematic b/o poor models, incomplete inhibition of target and poor reproducibility)
  - Phase I and II trials - is it inhibiting its target in patients? Is there a specific target *(biomarker)*? fundamental questions but undertaken rarely
  - Phase III trials - especially if biomarker defines a responsive population (e.g. *BRAF*, HER2)
Scientists should be encouraged to present whole data sets, not only positive results.
• We searched *major journals* Jan 2010 – Dec 2014 for RCTs which evaluated targeted agents for advanced solid tumours
• We searched references (& PubMed/Google) for supporting evidence from preclinical & early phase clinical studies.
• We characterized positive and negative phase III trials for:
  – Inclusion or not of a biomarker
  – Inclusion of translational/correlational research
  – Supporting evidence from early phase study

May-25-17
Success of phase III trials with and without correlative analysis

<table>
<thead>
<tr>
<th></th>
<th>Negative RCTs (N=60)</th>
<th>Positive RCTs (N=52)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase III trials with supporting correlative analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- In phase II studies</td>
<td>17 (28%)</td>
<td>33 (63%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>- In phase I studies</td>
<td>8 (13%)</td>
<td>17 (33%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Population enriched by a biomarker</td>
<td>8 (13%)</td>
<td>23 (44%)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Prior evidence for target inhibition in patients was rare for both negative and positive trials.
Survey of 55 Negative phase 3 trials

- 48/55 (87%) sponsored at least in part by Pharma
- Most common type of agent was anti-angiogenic
- No phase 2 trial identified prior to 20 (36%) phase 3 trials
- Minimal or no supporting evidence from phase 2 trials to support majority of negative phase 3 trials
  - Phase 2 data not done or not available at initiation
  - Different settings or comparator in phase 2
  - Metric of success not defined or met in phase 2
- No phase 2 studies evaluated target inhibition
Nine large RCTs of docetaxel+/- targeted agent for prostate cancer with ~9,000 patients at a cost of ~$1 billion

<table>
<thead>
<tr>
<th>Trial</th>
<th>Partner</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASCENT II</td>
<td>DN101 (calcitriol)</td>
<td>Poorer survival in experimental arm</td>
</tr>
<tr>
<td>VITAL II</td>
<td>GVAX vaccine</td>
<td>Poorer survival in experimental arm</td>
</tr>
<tr>
<td>SWOG S0421</td>
<td>Atrasentan</td>
<td>No difference in PFS or survival</td>
</tr>
<tr>
<td>ENTHUSE</td>
<td>Zibotentan</td>
<td>No significant difference in survival</td>
</tr>
<tr>
<td>MAINSAIL</td>
<td>Lenalidomide</td>
<td>No difference in survival (more toxicity)</td>
</tr>
<tr>
<td>CALGB 90401</td>
<td>Bevacizumab</td>
<td>No difference in survival (better PFS)</td>
</tr>
<tr>
<td>READY</td>
<td>Dasatinib</td>
<td>No difference in survival</td>
</tr>
<tr>
<td>VENICE</td>
<td>Aflibercept</td>
<td>No difference in survival (more toxicity)</td>
</tr>
<tr>
<td>SYNERGY</td>
<td>Custirsen (OGX-011)</td>
<td>No difference in survival</td>
</tr>
</tbody>
</table>
What can we learn from this dismal experience?

- Many of these trials (including VENICE) were undertaken with:
  1. Minimal evidence of effectiveness in preclinical models
  2. No evidence that the drugs inhibited their target in patients
  3. Lack of data in phase I and phase II trials

VENICE was initiated rapidly by Pharma because it was expected that docetaxel + bevacizumab would become the new standard treatment May 25-17.
(Lack of) appropriate early drug development:

- Poor reproducibility in preclinical experiments
- Few studies of target inhibition in patients
- Minimal or no evidence of benefit in phase II
3. RCTs may use surrogate endpoints that do not reflect patient benefit
Endpoints of clinical trials were discussed in my previous lecture

Main points:

1. Ultimate goals of any phase III trial are to improve the duration and/or quality of survival (OS or QoL)
2. Endpoints such as PFS or DFS are often poor surrogates for OS and QoL. Good correlation does not prove surrogacy.
3. Informative censoring (patients going off study prior to progression) leads to bias in PFS
4. Evaluation of Quality of Life (QoL) and Patient-reported outcomes (PROs) is done rarely and often poorly
5. Toxicity endpoints are under-reported

May-25-17
Questions (3)

About endpoints in phase 3 Randomized Controlled Trials
Which factors might limit relevance of RCTs?

4. Poorly-designed RCTs may show statistically significant but clinically irrelevant results

How can we evaluate clinical benefit or “value”
While FDA and EMA register drugs on basis of $p < 0.05$, Pharma will do large trials to detect small differences. Statistical significance ≠ clinical significance

Addition of Aflibercept to Fluorouracil, Leucovorin, and Irinotecan Improves Survival in a Phase III Randomized Trial in Patients With Metastatic Colorectal Cancer

 Previosly Treated With an Oxaliplatin-Based Regimen

Eric Van Cutsem, Josep Tabernero, Radek Lacko
Guy A. van Hazel, Vladimir Moiseenko, David Rémi Castan, and Carmen Allegra

< 2 months gain in median OS for a drug that is toxic and very expensive
Recent statement by the American Statistical Association about p-values

- P-values do not measure the probability that a hypothesis is true.
- Scientific conclusions and policy decisions should not be based only on p <0.05.
- A p-value does not measure the size of an effect or the importance of a result.
- By itself, a p-value does not provide evidence regarding a model or hypothesis.
The important test of validity of a clinical trial is not the p-value, it is whether the result is reproducible.
time-tested statistical methodologies are being modified, with excitement generated by meaningless gains in progression-free survival of a few weeks. A 2–3% improvement in overall survival seems to have become the outer limit of our intellectual expectation.
ESMO and ASCO scales for measuring clinical value

A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS)

Annals of Oncology 26: 1547–1573, 2015

& M. J. Piccart8,9

American Society of Clinical Oncology Statement: A Conceptual Framework to Assess the Value of Cancer Treatment Options

ESMO clinical value scale

ESMO MCBS evaluation

Curative

Curative-Evaluation form 1: for new approaches to adjuvant therapy or new potentially curative therapies

Non-curative

Non-curative-Evaluation forms 2a, b or c: for therapies that are not likely to be curative
Do contemporary drugs meet ESMO thresholds for meaningful clinical benefit?
Del Paggio et al; Annals Oncol (in press)

- 277 RCTS for breast, NSCLC, colorectal and pancreas cancer published 2011-2015
- 142 (51%) were positive for primary time-to-event endpoint
- Only 43/142 (30%) met ESMO threshold for clinical benefit
- Only 31% of RCTs were designed to detect or exclude differences in outcome that met the ESMO threshold for meaningful clinical benefit

We are spending our €€€ on ineffective treatments

May-25-17
Questions (4)

Evaluation of clinical benefit of new treatments
5. RCTs may be analyzed and reported prematurely...

... and mature results (when most patients have progressed or died) may never be published.
Premature reporting of clinical trials and failure to provide mature results

- RCTs are reported when a given number of “events” has occurred (often PFS events).
- Drugs may be registered by FDA/EMA based on very short F/U - look at the numbers under the S-curves.
- Results get less impressive with time.

<table>
<thead>
<tr>
<th>Median</th>
<th>First report</th>
<th>Updated</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR by Outcome</td>
<td>0.75</td>
<td>0.83</td>
<td>0.003</td>
</tr>
<tr>
<td>Grade 3-4 Adverse Events</td>
<td>8.5%</td>
<td>9.5%</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Chemohormonal Therapy in Metastatic Hormone-Sensitive Prostate Cancer

Christopher J. Sweeney, M.B., B.S., Yu-Hui Chen, M.S., M.P.H.

Hazard ratio for death with ADT+docetaxel, 0.61 (95% CI, 0.47–0.80) P<0.001

ADT+docetaxel (median overall survival, 57.6 mo)

ADT alone (median overall survival, 44.0 mo)

No. at Risk
ADT+docetaxel  397  333  189
ADT alone     393  318  168

Months

Median follow-up = 29 months
Everolimus in Postmenopausal Hormone-Receptor–Positive Advanced Breast Cancer

Median time on treatment ~15 weeks

OS (39 months F/U) Piccart et al, AO, 2014

HR = 0.89 (95% CI = 0.73–1.10)
Log-rank $P = 0.1426$
Kaplan–Meier medians
EVE + EXE: 31.0 months
PBO + EXE: 26.6 months

No. at Risk

<table>
<thead>
<tr>
<th>Everolimus</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVE + EXE</td>
<td>485</td>
</tr>
<tr>
<td>PBO + EXE</td>
<td>471</td>
</tr>
<tr>
<td>448</td>
<td>94</td>
</tr>
<tr>
<td>429</td>
<td>55</td>
</tr>
<tr>
<td>414</td>
<td>33</td>
</tr>
<tr>
<td>399</td>
<td>20</td>
</tr>
<tr>
<td>373</td>
<td>11</td>
</tr>
<tr>
<td>347</td>
<td>6</td>
</tr>
<tr>
<td>330</td>
<td>3</td>
</tr>
<tr>
<td>311</td>
<td>1</td>
</tr>
<tr>
<td>292</td>
<td>0</td>
</tr>
<tr>
<td>279</td>
<td>0</td>
</tr>
<tr>
<td>266</td>
<td>0</td>
</tr>
<tr>
<td>248</td>
<td>0</td>
</tr>
<tr>
<td>232</td>
<td>0</td>
</tr>
<tr>
<td>216</td>
<td>0</td>
</tr>
<tr>
<td>196</td>
<td>0</td>
</tr>
<tr>
<td>154</td>
<td>0</td>
</tr>
<tr>
<td>118</td>
<td>0</td>
</tr>
<tr>
<td>91</td>
<td>0</td>
</tr>
<tr>
<td>58</td>
<td>0</td>
</tr>
<tr>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Premature reporting of clinical trials and failure to publish mature results...

... uncertainty in the tails of survival curves
6. RCTs may underestimate toxicity of new treatments
All targeted agents add toxicity

There is substantial under-reporting of toxicity

- 58% of potentially fatal adverse events are not in the initial FDA drug label, and 39% are not reported in any published randomized trial

Toxicity is likely to be higher when new drugs are given to a less selected population

May-25-17
Possible reasons for under-reporting of harm:

• Assessment by clinicians does not represent experience of patients.

• Harm might be detected but not reported appropriately or reporting influenced by sponsor.

• Short follow-up in trials might not allow detection of long-term toxicity.

Because of selection of good performance status patients in trials, greater toxicity is likely when treatment is used in everyday practice.

Ethan Basch, Bryce B. Reeve, Sandra A. Mitchell, Steven B. Clauser, Lori M. Minasian, Amylou C. Dueck,
Questions (6)

Reporting of toxicity...

...from clinician-based to patient-based assessment
7. RCTs may be subject to biased reporting
• 92 of 164 trials showed no significant difference in their primary endpoint

• In 59% the concluding statement of the abstract used secondary endpoints to suggest benefit

• A positive result for the primary endpoint was associated with under-reporting of toxicity

• We have shown similar findings in RCTs published recently in the 5 leading journals
Honorary and ghost authorship

Honorary authorship occurs if any of its authors did not meet ICMJE authorship criteria:

35% of ~200 articles published in leading journals had an honorary author - usually a ‘thought leader’ and often first author.

Ghost authorship occurs when an individual who has substantial involvement in the study is not listed as an author.

At least 66% of 61 evaluable articles had an investigator or statistician as a ghost author, and ≥45% of 200 studies used a medical writer.

Medical writers are usually paid by companies – this may or may not lead to bias.
Editors of journals and reviewers should have a check-list to ensure:

1. Conclusions are based only on the primary endpoint.
2. Comprehensive reporting of toxicity.
3. Authorship is deserved and appropriate.
Authorship and bias in reporting trials
7. RCTs may select patients who do not represent those seen in every-day practice.
Efficacy = difference in outcome in an ideal population such as those selected to take part in a RCT

Effectiveness = difference in outcome in the real world – evaluated by health outcomes research applied to a population

A small difference in outcome in an RCT (i.e. in efficacy) with an agent that adds toxicity is likely to disappear when applied in routine practice (it may not be effective)
Men receiving 3-weekly docetaxel for CRPC | Routine practice at PMH | On-trial patients at PMH | P-value | TAX-327 trial
--- | --- | --- | --- | ---
Number | 314 | 43 | | 335 |
Median survival (months) | 13.6 | 20.4 | 0.007 | 19.3 |
% septic neutropenia | 9.6% | 0% | <0.001 | 3%
<table>
<thead>
<tr>
<th>RCTs</th>
<th>Population-based studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precise measures of efficacy under ideal conditions</td>
<td>Difficulty in eliminating bias and confounders of effect</td>
</tr>
<tr>
<td>Poor measure of effectiveness under real life conditions</td>
<td>Can estimate effectiveness in the general population</td>
</tr>
<tr>
<td>Limited information on toxicity</td>
<td>Assess toxicity under real life conditions</td>
</tr>
<tr>
<td>Applicability to clinical practice can be limited</td>
<td>Evaluate uptake of treatment in general population</td>
</tr>
</tbody>
</table>
Two RCTs show survival benefit for neo-adjuvant chemotherapy prior to cystectomy or radiotherapy – but patients are not often referred before surgery.

No single RCT has shown a survival advantage for adjuvant chemotherapy.

There is a Cochrane-based meta-analysis suggesting survival benefit but some included trials were of poor quality.
Adjuvant studies: (491 pts from 6 trials evaluating cisplatin-based combination chemotherapy

HR=0.75 (0.60-0.96)
P=0.019

Note uncertainty due to small sample
Immediate versus deferred chemotherapy after radical cystectomy in patients with pT3–pT4 or N+ M0 urothelial carcinoma of the bladder (EORTC 30994): an intergroup, open-label, randomised phase 3 trial

Cora N Sternberg, Iwona Skonczna, J Martijn Kerst, Peter Albers, Sophie D Fossa, Mads Agerbaek, Herlinde Dumez, Maria de Santis,

Lancet Oncol 2015; 16: 76–86

HR 0.54 (95% CI 0.40–0.73); p=0.00002

HR 0.78 (adjusted 95% CI 0.56–1.08); p=0.13
A key question is whether one will see benefits in patients treated in the community.
Questions:

1. What is the uptake in Ontario of Neoadjuvant (NACT) and Adjuvant chemotherapy (ACT)?

2. Does it improve survival in the community?


5000+ path reports obtained to assign stage.
A propensity score allows analysis of observational studies so that it mimics characteristics of an RCT.

It is a balancing score – conditional on the propensity score the distribution of baseline factors will be similar between treated and untreated subjects.
Results of Health Outcomes Study

2944 cases

- 5-year overall survival = 29% (lower than in clinical trials)
- Use of perioperative chemotherapy (using propensity score to correct for confounding factors) associated with improved survival:
  \[ HR = 0.71 \text{ (95\% CI 0.62-0.81), } p <0.001 \]
- Similar effect on cancer-specific survival

May-25-17
Adjuvant chemotherapy for bladder cancer?

<table>
<thead>
<tr>
<th>Citation</th>
<th>Study design</th>
<th>Sample size</th>
<th>Hazard Ratio/Relative risk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ACT/Total</td>
<td>OS/CSS</td>
</tr>
<tr>
<td>ABC 2005</td>
<td>Individual patient data meta-analysis from 6 RCTs</td>
<td>246/491</td>
<td>0.75</td>
</tr>
<tr>
<td>Svatek 2010</td>
<td>Retrospective cohort study from 11 high volume centers</td>
<td>932/3947</td>
<td>0.83</td>
</tr>
<tr>
<td>Leow 2014</td>
<td>Literature-based meta-analysis from 9 RCTs</td>
<td>475/945</td>
<td>0.77</td>
</tr>
<tr>
<td>Booth 2014</td>
<td>Population-based Retrospective cohort study</td>
<td>541/2809</td>
<td>0.71</td>
</tr>
<tr>
<td>Sternberg 2015</td>
<td>Largest RCT (EORTC 30994)</td>
<td>141/248</td>
<td>0.78</td>
</tr>
</tbody>
</table>
Observational health outcomes studies vs. RCTs
Acknowledgements

My thanks to many present and former fellows, who undertook much of the work presented, and stimulated many of the ideas that led to it. Especially:

Eitan Amir
Chris Booth
Alberto Ocana
Bostjan Seruga
Arnoud Templeton
‘Paco’ Vera-Badillo
This talk was based on:

(i) An invited talk at the TATA Memorial Hospital Jubilee Conference in Mumbai, Feb 2016

(ii) The following paper that is in press in Lancet Oncology

“Tannock et al: The relevance of randomised controlled trials to clinical practice”
Endpoints for clinical trials in cancer patients

Ian F. Tannock

Princess Margaret Hospital and University of Toronto

May-25-17
Potential conflicts of interest

I have advised multiple companies about design of clinical trials for which I have received contributions to my research fund.

I have chaired international company-sponsored trials for hormone-refractory prostate cancer (TAX-327, VENICE)

I do not speak for companies and do not participate in Pharma-sponsored satellite symposia
Endpoints need to reflect the goals of different phase clinical trials

Phase I trials
- “Is the new treatment tolerable?”
- “What are the main toxicities?”
- “What is the recommended dose for phase II”

Phase II trials
- “Is there evidence of anti-tumour activity?”

Phase III trials
- “Is there evidence of patient benefit?”
Endpoints for Phase II trials

Main aim is to demonstrate (or exclude) anti-tumour activity

- Tumour response rate (RECIST criteria)
- Tumour marker response (e.g. PSA response)
- Progression-Free Survival (PFS is only interpretable in randomized phase II trials)

Important to recognize that demonstration of anti-tumour activity (e.g. tumour response) does not necessarily imply benefit to patients.
New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada.


New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1).

Main features of RECIST Criteria

- Define up to 5 target measurable lesions (M: >20mm or >10mm by spiral CT) and non-measurable (non-M) lesions
- CR requires disappearance of both M and non-M
- PR requires ≥30% shrinkage in sum of longest diameters of M without progression of non-M
- Progression requires ≥20% increase in sum of longest diameters of M, clear progression in non-M or new lesions
- Stable Disease (SD) if do not meet criteria of response or progression
- Response category requires confirmation after ≥4 weeks if RR is primary endpoint
RECIST criteria have been very useful in unifying criteria for evaluating tumour response. However:

• Several studies have shown poor inter-observer reproducibility. Estimates of PR are subject to error.
• Stable Disease (SD) is a signal to continue treatment, and does not necessarily imply an effect of treatment.
• Errors persist when tumour response is shown in “waterfall plots”
8891 patients randomized to receive a placebo or best supportive care (BSC) in 67 RCTs

• SD reported in 53% of these trials

• Median rate of SD=25% (range 0-67%)

Many tumours growing steadily will satisfy criterion of short-term SD

SD >6 months might be meaningful
Waterfall and Swimmer Plots
(From Brana et al. Invest New Drugs 2014;32:1269-77)
Progression (+20%)

PR (-30%)

Different investigators will record between 1 and 5 patients as having PR
Many authors report a rate of “Clinical Benefit” defined as % of patients with CR + PR + SD.

This definition does NOT imply benefit to patients:

“Clinical Benefit” ≠ Clinical Benefit
Questions?

• Defining tumour response

• Endpoints for phase II trials
Phase III Trials
Which 1° endpoint should be used in a phase III RCT?

There are only two goals of any new treatment:

To allow the patient to live longer

and/or

To allow the patient to live better

Hence, there are only two important endpoints of a phase III trial:

1. Overall Survival
2. Quality of Survival

Anything else is a surrogate endpoint
Survival and its quality are the endpoints of clinical trials that evaluate survivorship.

Surrogate endpoints

- Many phase III trials have a $1^\text{st}$ outcome measure other than OS or QoL.
- DFS and PFS are commonly used in trials of systemic therapy.
- They have been used for registration of new drugs but are often poorly correlated with OS.
- They are also subject to considerable bias.

May-25-17
There is perfect correlation between true and proposed surrogate for each group - but: experimental group (E) has larger true endpoint than controls (C) but smaller potential surrogate endpoint.
PFS is subject to biased reporting

Due to:

1. Different times of evaluation (especially in unblinded trials)

2. If there is uneven drop-out of patients prior to progression in the experimental and control arms
The Clinical Viewpoint: Definitions, Limitations of RECIST, Practical Considerations of Measurement

Liza C. Villaruz and Mark A. Socinski


A) Prespecified evaluation interval
   Disease progression is reported at the pre-specified evaluation time

B) Evaluation bias
   An evaluation might be done out of turn, as might be the case if symptoms prompt concern for disease progression

C) Measurement bias
   Toxicity and treatment delays might delay an evaluation

D) Attrition bias
   Patients lost to follow-up may not have the same risk trajectory as patients remaining on trial

E) Informative censoring
   Central review might have determined disease progression after local determination of disease progression resulting in missing data

F) Interval censoring
   A patient’s true date of progression likely lies between scheduled disease evaluations

Evaluation for disease progression → Documentation of disease progression (local review)
True date of disease progression → Documentation of disease progression (central review)
An example: Should we adopt exemestane + everolimus as 2\textsuperscript{nd} line treatment for post-menopausal women with ER+ breast cancer?
Everolimus in Postmenopausal Hormone-Receptor–Positive Advanced Breast Cancer

José Baselga, M.D., Ph.D., Mario Campone, M.D., Ph.D..


**PFS**

Hazard ratio, 0.43 (95% CI, 0.35–0.54)

P<0.001 by log-rank test

**Serious Adverse Events (SAEs) due to treatment:**

11% vs. 1%

Early stopping due to SAE or withdrawn consent:

24% vs 6%

This trial led to registration of everolimus + exemestane for postmenopausal ER+ women ($$$$

May-25-17
Censoring bias: Patients who withdraw from a study (for toxicity or other reasons) are censored if they have not satisfied criteria of progression. A better representation might be difference in “time on treatment” and survival results are now available (Piccart et al, Annals Oncol, 2014)
• Trials of bevacizumab + CT vs CT alone for Ca breast, prostate, lung, pancreas, stomach, ovary
  – Are reported as negative if primary endpoint is OS
  – Are reported as positive if primary endpoint is PFS
• In 2 trials the company changed the endpoint from OS to PFS while the trial was in progress
  – In all trials bevacizumab added toxicity
What should we call an agent that increases PFS, has no effect on survival, and adds toxicity?

HARMFUL
Primary endpoints in phase 3 trials should be overall survival or its quality (or valid surrogates)

We should be critical of phase 3 trials for people with incurable cancer that fail to measure quality of survival
QoL as an essential component of Survivorship

FDA and EMA have been reluctant to use QoL or PROs as endpoints for drug registration.

Drugs should be registered on basis of benefit > risk

In trials designed to provide palliation (most of them) – QoL or a PRO should be a co-primary endpoint...

... they are major determinants of benefit to risk ratio.

Toxicity is under-reported in trials and the new PRO-CTCAE might help to change the balance
Quality of Life (QoL) and Patient-Reported Outcomes (PROs)

Many trials are directed to palliation of subjects with incurable cancer, but few of them include QoL or PROs as primary outcomes.

RCTs that include QoL often assess it badly: they measure "average QoL" in each arm at baseline and some time later --- like measuring "average tumour size" in each arm instead of response rate.

QoL or PROs related to a major symptom (e.g. pain) can be measured reliably and objectively. This should be recorded as % of patients having a defined improvement.

There are several well validated questionnaires. Most useful are the FACT and EORTC questionnaires.

"THE QUALITY, NOT THE LONGEVITY, OF ONE'S LIFE IS WHAT IS IMPORTANT."

- MARTIN LUTHER KING, JR.
We are interested in some things about you and your health. Please answer all the questions by circling the number that best applies to you. There are no “right” or “wrong” answers. The information that you provide will remain strictly confidential.

Please fill in your initials: .................................................................

Your birthdate (Day, Month, Year): ...................................................

Today's date (Day, Month, Year): ......................................................

<table>
<thead>
<tr>
<th>Question</th>
<th>NO</th>
<th>YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Do you have any trouble doing strenuous activities, like carrying a</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>heavy shopping bag or a suitcase?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Do you have any trouble taking a long walk?</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3. Do you have any trouble taking a short walk outside of the house?</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4. Do you have to stay in a bed or a chair for most of the day?</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5. Do you need help with eating, dressing, washing yourself or using the</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>toilet?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Are you limited in any way in doing either your work or doing household jobs?</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7. Are you completely unable to work at a job or to do household jobs?</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
29. How would you rate your overall physical condition during the past week?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Very poor</td>
<td>Excellent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>¿Cómo valoraría su condición física general durante la semana pasada?</th>
<th>Pésima</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>¿Cómo valoraría su calidad de vida general durante la semana pasada?</th>
<th>Pésima</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
FACT-G (Version 4)

Below is a list of statements that other people with your illness have said are important.

<table>
<thead>
<tr>
<th>PHYSICAL WELL-BEING</th>
<th>Not at all</th>
<th>A little bit</th>
<th>Somewhat</th>
<th>Quite a bit</th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>I have a lack of energy</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>I have nausea</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Because of my physical condition, I have trouble meeting the needs of my family.</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>I have pain</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>I am bothered by side effects of treatment</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>I feel ill</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>I am forced to spend time in bed</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Example of a registration trial that used pain relief as a primary endpoint

Chemotherapy With Mitoxantrone Plus Prednisone or Prednisone Alone for Symptomatic Hormone-Resistant Prostate Cancer: A Canadian Randomized Trial With Palliative End Points

By Ian F. Tannock, David Osoba, Martin R. Stockler, D. Scott Ernst, Alan J. Neville, Malcolm J. Moore, George R. Armitage, Jonathan J. Wilson, Peter M. Venner, Christopher M.L. Coppin, and Kevin C. Murphy

**Conclusion**: Chemotherapy with mitoxantrone and prednisone provides palliation for some patients with symptomatic hormone-resistant prostate cancer.
Docetaxel and prednisone has been standard 1st-line chemotherapy since 2004.
TAX-327: Secondary endpoints

<table>
<thead>
<tr>
<th>TAX-327: 2° endpoints</th>
<th>Docetaxel q 3wk</th>
<th>Mitoxantrone q 3wk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain Response Rate</td>
<td>34.6% p=0.01</td>
<td>21.7%</td>
</tr>
<tr>
<td>PSA Response Rate</td>
<td>45.4% p=0.0005</td>
<td>31.7%</td>
</tr>
<tr>
<td>QOL Response rate</td>
<td>21.9% p=0.009</td>
<td>13.1%</td>
</tr>
</tbody>
</table>

Which is more important? 3-month improvement in survival? or Improved pain and quality of life?
• 92 of trials published 1995-2011 showed no significant difference.

• In 59% of reports of these 92 trials the concluding statement of the abstract used secondary endpoints to suggest benefit.

• Only one third of reports indicated the frequency of grade 3-4 toxicity in the abstract: under-reporting of toxicity associated with positive primary outcome.

• Similar results from a larger study of trials for 4 disease sites in 5 leading journals.
Questions?

- Endpoints for phase III trials
- Surrogate endpoints
- Quality of Life and Patient-reported Outcomes