CRITICAL APPRAISAL OF A MEDICAL PUBLICATION OR PRESENTATION

How to evaluate reports of Phase 3 trials

Ian F Tannock MD, PhD, DSc
Emeritus Professor of Medical Oncology and Medical Biophysics
Princess Margaret Cancer Centre and University of Toronto
EXCELLENT PRACTICE-CHANGING TRIALS THAT COMPARED DIFFERENT STRATEGIES INCLUDE…

- PROTEC-T trial comparing surgery, radiotherapy and watchful waiting for screen-detected prostate cancer
- STAMPEDE trial adding new therapies (abiraterone, docetaxel, radiotherapy) to standard hormonal treatment for advanced prostate cancer
- Trials of immunotherapy for melanoma, lung cancer and kidney cancer

Some other trials have not been designed, analysed and reported with the same rigour.
QUESTIONS TO ASK ABOUT PHASE 3 CLINICAL TRIALS

Is the control group appropriate?
Are the outcome measures (endpoints) appropriate?
Is toxicity evaluated comprehensively?
Are the outcomes of clinical value, as well as statistically significant? Do related trials give consistent results?
Does the report of the study reflect its results?
Do the results apply to patients I see in my clinic?

Relevance of randomised controlled trials in oncology

A PHASE III RANDOMISED TRIAL SHOULD...

.... Compare an experimental treatment with an appropriate standard of care
If oncologists accept different standards of care, then control groups can include more than one option

Example: Pembrolizumab vs. 2nd line chemotherapy for advanced bladder.
Docetaxel, paclitaxel or vinflunine were options for control treatment
Some trials have “loaded the dice” with control treatment **NOT** a current optimal standard of care

Examples:
Radiotherapy + cetuximab vs. radiotherapy alone for locally advanced head and neck cancer¹
It was known already that radiotherapy + concurrent cisplatin improved survival
PROFOUND: Olaparib vs. enzalutamide (after abiraterone) or vice versa for castrate-resistant prostate cancer (CRPC) with DNA repair defects² ³
35% had not received docetaxel, and 80% had not received cabazitaxel, both known to improve survival for men with CRPC


A PHASE III RANDOMISED TRIAL SHOULD HAVE AN APPROPRIATE PRIMARY ENDPOINT

Outcome measure

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 **ultimate** endpoints of a phase III trial:

1. Overall Survival (OS)

2. Quality of Survival (QoL)

Other endpoints should be shown to be surrogates for OS or QoL
MANY ANTI-CANCER DRUGS APPROVED BY EMA & FDA HAVE NOT IMPROVED DURATION OR QUALITY OF SURVIVAL

48 drugs approved by EMA for 68 indications in 2009–2013

At time of approval:

- Improvement in survival for 24 indications (35%)
- Improved survival by 1–6 months (median 2.7 months)
- Improved quality of life (QoL) for 7 indications (10%)

With 5-year follow-up 35/68 drugs (51%) showed improvement in survival and/or QoL

Similar findings for FDA

HOW TO COMPARE SURVIVAL CURVES?

Hazard Ratio (HR) = ratio of events of experimental to control arm in any time interval
Assumes proportional hazards (constant ratio of events over time)
HR is quoted but invalid for many trials (crossing curves, plateaus…)
e.g.


Need alternative analyses, such as Biomarker-separated curves, Plateau survival values or Restricted mean survival time.
QUALITY OF LIFE AS AN ENDPOINT IN CLINICAL TRIALS

The goal of many trials is “palliation”…
…but few trials include QoL or a patient-reported outcome (PRO) as a major endpoint
… and when included they often use flawed analysis
In most RCTs, investigators measure…
…a mean QoL score for each group at baseline & at some later time.
They compare mean scores at that time (e.g. 3 months)…
....for each arm with its baseline and between arms

“Mean QoL” is a meaningless concept!
QoL is a property of an individual that can improve or get worse

QL must be evaluated rigorously. It is important to:
1. Select one aspect of QL as the primary endpoint
2. Establish a hypothesis about the magnitude of change in that endpoint that is clinically important
3. Evaluate changes in other aspects of QL as supportive information

Analysis of multiple endpoints without an a priori hypothesis will lead to artefacts
PREFERRED METHOD FOR ASSESSING QOL OR SYMPTOM CONTROL IN A CLINICAL TRIAL

Select a relevant measure of QoL (or dominant symptom) and measure it in each patient at baseline

Specify improvement in that measure for a QoL response

Measure QoL of each patient repeatedly

Determine the proportion of patients in each arm who satisfy the predetermined criterion of response

Determine median and range of duration for the QoL response in each arm

Learn from assessment of Tumour Response: we don’t average tumour size among patients
POSSIBLE SURROGATE ENDPOINTS

Many phase III trials have a 1° outcome measure other than Overall Survival (OS) or Quality of Life (QoL)

Disease-Free Survival (DFS) is used often in adjuvant trials

Progression-free Survival (PFS) is used in trials of systemic therapy for advanced disease

Local control is used in trials of radiotherapy or surgery

Many assume that these endpoints convey benefit, often not supported by evidence

ADDING DRUGS TO ADJUVANT THERAPY MAY IMPROVE DFS BUT NOT OS

Applying treatment at relapse might maintain equal survival and avoid added toxicity and cost from treating all patients at risk

Relationship between 2-year DFS and 5-year OS in adjuvant trials for breast cancer

N.B. Improvements in DFS predict smaller (or no) improvements in OS


© 2008 European Society for Medical Oncology. Published by Elsevier Ltd. All rights reserved.
SHOULD WE ADD A CDK4/6 INHIBITOR TO ADJUVANT THERAPY FOR BREAST CANCER?

CDK4/6 inhibitors ($$$$) improve PFS (& probably OS) in women with HR+ metastatic breast cancer
MonarchE trial showed improved invasive DFS when abemaciclib added to adjuvant therapy

This should not be a new standard.
Also, PALLAS trial of adjuvant palbociclib is negative.

N.B.
Very few patients at risk define tails of curves
PROGRESSION-FREE SURVIVAL

PFS is used commonly as a primary endpoint of clinical trials

Rarely a surrogate for overall survival

Value of PFS depends on balance between:
   A. added toxicity
   B. improvement or delay in symptoms

Improvement in symptoms cannot be assumed to dominate added toxicity (Is there better QoL?)

PFS is subject to bias when there is uneven dropout between the arms of a trial (known as Informative censoring)
EVEROLIMUS IN POSTMENOPAUSAL HORMONE-RECEPTOR-POSITIVE ADVANCED BREAST CANCER


Early stopping due to SAE or withdrawn consent: 24% vs. 6%
Patients withdrawn before progression were censored

This result led to registration of everolimus + exemestane for postmenopausal ER+ women (↑toxicity & $$$$)

PFS Hazard ratio, 0.43 (95% CI, 0.35–0.54) P<0.001 by log-rank test

Everolimus plus exemestane (median PFS, 6.9 mo)
Placebo plus exemestane (median PFS, 2.8 mo)
INFORMATIVE CENSORING – A NEGLECTED CAUSE OF BIAS IN ONCOLOGY TRIALS

BOLERO-2: Time on treatment
(courtesy of Greg Pond)

... and survival results???

HR=0.89, P=0.14

AN AGENT THAT INCREASES PFS...

... has no effect on survival, does not improve QoL, and adds toxicity is HARMFUL

Alternative definitions of PFS?
USE CAUTION WHEN EVALUATING OUTCOMES FOR RARE TUMOURS AND BIOMARKER-DEFINED SUBGROUPS

Basket trials: Testing an agent targeting a mutation in various cancer types

Umbrella trials: Testing agents targeting different mutations in a single cancer type

Response rate (usual endpoint) does not ensure patient benefit

Beware the shrinking denominator

1. Chan & Tannock IF. J Clin Oncol 2021

Based on >20,000 patients in basket trials

1000\textsuperscript{a} patients undergo genetic testing

800\textsuperscript{b} have genetic profile

311\textsuperscript{b} have a targetable mutation

114\textsuperscript{b} receive a matched drug

1-30\textsuperscript{c} respond

\textsuperscript{1} Based on >20,000 patients in basket trials
A PHASE 3 RANDOMISED TRIAL SHOULD EVALUATE TOXICITY COMPREHENSIVELY

All targeted agents add toxicity. Patients experience more toxicity than reported by doctors.

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 RCT.

Toxicity is higher when drugs are given to patients in the real world.

Under-reporting of harm in clinical trials

THE MISSING VOICE OF PATIENTS IN DRUG-SAFETY REPORTING


© 2012 by American Society of Clinical Oncology.
STATEMENT BY THE AMERICAN STATISTICAL ASSOCIATION ABOUT P-VALUES

P-values do not measure the probability that a hypothesis is true

Scientific conclusions & 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

EMA /FDA and major journals have ignored this message. Oncologists should lobby for clinical value, (not statistical significance) to be used as the criterion of importance
CLINICAL VALUE SCALES

ESMO (and ASCO and NCCN) have scales to assess magnitude of clinical benefit (MCBS) or “value”
They emphasise improvements in duration or quality of survival

Consistency among trials is also very important

MANY RCTS DO NOT MEET ESMO CRITERIA FOR MEANINGFUL CLINICAL BENEFIT

277 RCTS for breast, NSCLC, colorectal and pancreas cancer published 2011-2015
51% were positive for primary time-to-event endpoint
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

- 65% of trials with curative intent
- 18% of trials with palliative intent

If a trial is reported as positive, ask whether the experimental treatment meets criteria of clinical benefit

DOES THE REPORT OF A STUDY REFLECT ITS RESULTS?

In reports of phase 3 RCTs…

1. The primary endpoint should be defined explicitly in the abstract of the paper or presentation
2. That primary endpoint should reflect benefit to patients
3. The concluding sentence of the abstract should relate only to the primary endpoint
4. The abstract should include a statement of major toxicity

Many reports, both in high level journals and at ESMO and ASCO meetings, do NOT meet these simple requirements
BIAS IN REPORTING OF END POINTS OF EFFICACY AND TOXICITY

In randomised, clinical trials for women with breast cancer

92 of 164 trials published 1995-2011 showed no significant difference in 1° endpoint

In 59% of reports of these 92 trials, concluding statement of abstract used 2° endpoints to suggest benefit

One third of abstracts indicated frequency of grade 3-4 toxicity

Honorary and Ghost Authorship was common

Similar findings in RCTs published for other malignancies in 5 leading journals

RCTs are usually powered to analyse a 1º endpoint for the whole group. Secondary & Subgroup analysis can lead to false results

e.g., Adjuvant 5-FU and levamisole for Dukes C colon cancer:

1. Mayo Clinic Trial\(^1\)
   More effective for men, older patients

2. Southwest Oncology Group Trial\(^2\)
   More effective for women, younger patients

GHOST WRITING

80% of 2020 ESMO abstracts reporting Pharma-sponsored Phase 3 trials acknowledged “writing assistance”

16/21 reports of 3 months publications of such trials in NEJM were first drafted by a medical writer

While some medical writers might draft unbiased reports, they are employed by or contracted to the sponsor

There is potential for biased reporting
THE SPIIN TRIAL: EFFECT OF BIASED REPORTING IN ABSTRACTS OF RCTs FOR CANCER

For abstracts with spin, the experimental treatment was rated to be more beneficial.

Logged on to the system (N=433)

Randomly allocated (N=300)

Excluded:
Not clinicians (n=122)
Logged on to the system but did not evaluate an abstract (n=11)

Allocated to assess one abstract among the 30 abstracts with spin (n=150)

Included in final analysis (n=150)

Allocated to assess one abstract among the 30 abstracts without spin (n=150)

Included in final analysis (n=150)

THE ROLE OF EDITORS AND REVIEWERS

Reporting of clinical trials can be improved if:

- Reviewers are provided with a checklist of simple questions
- Editors verify that the questions are addressed – at least for the abstract

The checklist should ask:

- Is the primary endpoint appropriate and defined explicitly?
- Does the concluding sentence of the abstract and the main conclusion of the paper apply only to that primary endpoint?
- Does the abstract and paper describe toxicity adequately?
- If subgroups are analysed, were they pre-planned and exploratory?
- Is the report free of spin and bias?
WILL THE RESULTS OF THE CLINICAL TRIAL APPLY TO PATIENTS IN MY CLINIC?

The Efficacy–Effectiveness Gap

Patients are highly selected to take part in clinical trials (younger, high PS, comorbidity excluded)

Efficacy = Benefit of a new treatment in clinical trials

Effectiveness = Benefit of a new treatment in real world

2. Image reproduced by permission of Malcolm Willett.
TRANSLATING TRIALS TO CLINICAL PRACTICE

Docetaxel for castrate-resistant prostate cancer (CRPC)

<table>
<thead>
<tr>
<th>Men receiving 3-weekly docetaxel for CRPC</th>
<th>Routine practice at PMCC</th>
<th>On-trial patients at PMCC</th>
<th>P-value</th>
<th>TAX-327 trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>314</td>
<td>43</td>
<td></td>
<td>335</td>
</tr>
<tr>
<td>Median survival (months)</td>
<td>13.6</td>
<td>20.4</td>
<td>0.007</td>
<td>19.3</td>
</tr>
<tr>
<td>% septic neutropenia</td>
<td>9.6%</td>
<td>0%</td>
<td>&lt;0.001</td>
<td>3%</td>
</tr>
</tbody>
</table>

Another Example: 5-year OS after radical cystectomy for muscle-invasive bladder cancer:
In clinical trials: ~50%; In Ontario Cancer Registry (N~3000): ~30%

ASSESSMENT OF OUTCOMES
Associated with the use of newly approved oncology drugs in medicine beneficiaries

Compared >11,000 trial participants with >9,000 Medicare patients

Four examples of many treatments:

Medicare patients treated with FDA-approved drugs did not live as long as trial participants and had more dose reductions.
RANDOMISED CONTROLLED TRIALS AND POPULATION-BASED OBSERVATIONAL RESEARCH

Partners in the evolution of medical evidence

<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>

RCTS AND REAL-WORLD STUDIES CAN BE COMBINED TO PROVIDE CONVINCING EVIDENCE

Example: adjuvant chemotherapy for bladder cancer
Underpowered RCTs suggest survival benefit (HR ~0.76)

- Meta-analysis of 9 RCTs (N=945)\(^1\)
- EORTC study (N=284)\(^2\)

3 large well conducted real-world studies suggest survival benefit with HR = 0.83, 0.71, 0.70\(^3\)

Consistency of the evidence indicates a convincing role for cisplatin-based adjuvant chemotherapy

Most large RCTs are funded by commercial sponsors.

The goals of (i) generating profit & (ii) improving the outcome for cancer patients are sometimes, but not always, convergent.

**Oncologists should...**

... have a critical but constructive mindset when reading reports and observing presentations of results of clinical trials.

... lobby EMA and FDA to approve new treatments on the basis of clinical benefit rather than statistical significance.
THANK YOU!