Understanding the needs for statistical evidence of decision makers in medicine

What are the options, and who is asking the question?

Elizabeth Garrett-Mayer, PhD
Director of Biostatistics and Data Management
Center for Research and Analytics (CENTRA)
American Society of Clinical Oncology
Alexandria, VA

American Statistical Association's Symposium on Statistical Inference
October 11-12, 2017 • Bethesda, MD

ASCO

Background

- Anti-cancer drug development, early through late phase

- Context:
  - Incremental improvements in efficacy
  - Combinations of modestly effective agents compared to single agents
  - Range of toxicity effects
  - Huge range of costs to patients, with many costing 10's of thousands of $$ per month.

- What evidence do we need?
- *What decisions need to be made?*

ASCO

Statistical Evidence*

Imagine a randomized trial, comparing drug regimen A to drug regimen B, were just performed and the data have been analyzed.

Three relevant questions:

1. **What do I believe, now that I have observed the results?**
2. **What should I do, now that I have observed the results?**
3. **What do the results of this trial tell me about regimen A versus B?**

- Question 3 is critical to the reporting of statistical data in scientific journals.
- But, question 2 focuses on **decision making**.

The answer to question 3 informs our answer for question 2, but there is much more needed to address question 2.

ASCO

*Statistical Evidence: A Likelihood Paradigm. Richard Royall, 1997 (Chapman and Hall).

Different decision making scenarios

A. Do the results of my phase II trial of A vs. B provide sufficient evidence to justify proceeding to the phase III setting?
B. Do the results of this phase III trial provide sufficient evidence for the FDA to approve treatment regimen B?
C. Do the results of this phase III trial provide sufficient evidence for a doctor to prescribe regimen B instead of regimen A to her patient?
D. Do the results of this phase III trial provide sufficient evidence for single payer systems (e.g. Europe) to justify providing regimen B to its patients?
E. Do the results of this phase III trial provide evidence and information regarding who should receive regimen B (vs A?) in resource-poor settings?

ASCO

Different decision making scenarios

A. Do the results of my phase II trial of A vs. B provide sufficient evidence to justify proceeding to the phase III setting?
B. Do the results of this phase III trial provide sufficient evidence for the FDA to approve treatment regimen B?
C. Do the results of this phase III trial provide sufficient evidence for a doctor to prescribe regimen B instead of regimen A to her patient?
D. Do the results of this phase III trial provide sufficient evidence for single payer systems (e.g. Europe) to justify providing regimen B to its patients?
E. Do the results of this phase III trial provide evidence and information regarding who should receive regimen B (vs A?) in resource-poor settings?

ASCO

A. Phase II → Phase III?

- Phase II trials are generally designed to provide a ‘go / no-go’ decision at the conclusion.
- What is the trial design?
  - Single arm trial of regimen B alone?
  - Randomized A vs. B?
- **Strong emphasis on traditional statistical inference, such as p-values, for efficacy only endpoint.**
  - Many talks at this conference regarding alternatives.
- Toxicity is considered secondary, usually does not stop development at this stage.
- Some movement to designs that incorporate a bivariate endpoint of toxicity and efficacy.
- Cost of agent? Not formally considered in the design.

ASCO

A. Phase II → Phase III?

- Who is making the decision here? The drug developer.
- What determines movement to Phase III?
  - Efficacy 'signal' (and in which patients?)
  - Manageable toxicity
  - Cost?
- Cost benefit analysis
  - Likelihood of successful phase III trial→ likelihood of FDA approval
  - Cost of phase III program
  - Return on investment: how large is the patient population?
    - Rare cancers → low revenue
    - Common cancers → high revenue
- Decision making?
  - Phase II trial efficacy and toxicity results are key piece of the puzzle
  - They are used to inform cost benefit analysis.
  - Does it come down to a 'go vs. no-go' p-value? Usually not.
- *Statistical evidence is a combination of early phase trial results and forecasting financial rewards.*

ASCO

Different decision making scenarios

A. Do the results of my phase II trial of A vs. B provide sufficient evidence to justify proceeding to the phase III setting?
B. Do the results of this phase III trial provide sufficient evidence for the FDA to approve treatment regimen B?
C. Do the results of this phase III trial provide sufficient evidence for a doctor to prescribe regimen B instead of regimen A to her patient?
D. Do the results of this phase III trial provide sufficient evidence for single payer systems (e.g. Europe) to justify providing regimen B to its patients?
E. Do the results of this phase III trial provide evidence and information regarding who should receive regimen B (vs A?) in resource-poor settings?

ASCO

B. FDA Medical Decision Making

- FDA mission: FDA is responsible for protecting the public health by assuring the safety, efficacy and security of human and veterinary drugs, biological products, medical devices, our nation's food supply, cosmetics, and products that emit radiation.
- "We have to be sure we are not approving a placebo"—Rick Pazdur
- And, what about toxic placebos?
- FDA criteria
  - Strict alpha level: type I errors must be rare.
  - Relaxed beta level:
    - Onus is not on the FDA to ensure that effective drugs are brought to market.
    - Drug developers should ensure that trials are designed to detect a meaningful benefit.

ASCO

B. FDA Medical Decision Making

- Recent ODAC meeting evaluating Sutent for adjuvant kidney cancer (September 2017)
- What is ODAC? Oncologic Drugs Advisory Committee.
- Instructions to ODAC:
  - Do not consider the guidelines of professional groups in your vote,
  - View subset analyses from other trials with a grain of salt, and
  - If you have reasons to believe that DFS is the wrong endpoint in this setting, address this in your comments
- 12 member ODAC: 6-6 vote on Sutent.
  - Interpretation? “Maybe.”
  - Randomized trial of Sutent vs. placebo in patients with resected cancer
  - Primary endpoint: disease-free survival (DFS)

ASCO

* The Cancer Letter, Sept 22 2017, v 43, no 35. www.cancerletter.com

* https://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/OncologicDrugsAdvisoryCommittee/ucm576931.htm

B. FDA Medical Decision Making

- Question to ODAC: Is the benefit-risk profile of Sutent acceptable for the adjuvant treatment of patients at high risk of recurrent renal cell carcinoma following nephrectomy?

- Notes:
  - DFS hazard ratio: 0.76 (95% CI: 0.59, 0.97); p = 0.03.
  - OS hazard ratio (early): 0.92 (95% CI: 0.66, 1.28)
  - Grade 3-4 adverse events in 60% and 15% of patients in Sutent and placebo arms, respectively.

* The Cancer Letter, Sept 22 2017, v 43, no 35. www.cancerletter.com

* https://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/OncologicDrugsAdvisoryCommittee/ucm576931.htm

I voted Yes, and I want to tell a very brief story as to why I voted yes. About 17 years ago, my urologist in Kansas City called me and he said, "Dan, we got the scan back, and you have kidney cancer." And I was stunned. And I was so stunned that I said to him, I said, "Doc, is this going to get me? And he says, 'No, I don't think so.' He says, 'There's hope." So, when my wife came home I said, "I have kidney cancer and I'm thinking of Bill Clinton." She says, "Are you nuts? Why do you think of Bill Clinton?" I said, "Because he always said he was from a town named Hope. And I have hope that I'm going to survive this." So, I feel this drug provides patients with some hope.

Richard Lumley
Patient representative, Kansas City, Missouri

I voted Yes. This was one of the most difficult decisions. What really swayed me was that they do have a well-designed randomized study that shows DFS [magnitude] that was the basis for approval of many different agents in many different [indications]. Having seen the data we should at least offer for physicians and patients to at least discuss these data.

Ramaprasad Srinivasan
Investigator and head Molecular Cancer Therapeutics Section Urologic Oncology Branch, NCI Center for Cancer Research

I voted No. The problem I have goes back to sitting in front of a patient and being able to digest these data and helping them digest these data, telling them that this is going to delay the [recurrence] of their kidney cancer. I don't see data supporting this. I don't see data that this may delay the onset or that we are going to delay the time that you're going to need additional treatment, which is a relevant question, and I did not see data to support this.

Bruce Redman
Professor of medicine, Department of Internal Medicine Division of Hematology/Oncology, University of Michigan Comprehensive Cancer Center

I voted No, and similar to my peers I struggled with this.

Just to give you brief background on me, I am a statistician, who has been designing clinical trials in this space, so, of course, when I see something positive, it's exciting.

Similar to my peers, the thing that I was struggling most with was the robustness of data, if we can make a decision based on 257 events. I have no issue with disease-free survival as an endpoint in other diseases, colon and breast. In the adjuvant setting it has not been extensively used. But to base a decision on 257 events when the confidence interval almost touches 1 is troubling. And when you look at the toxicity profile, it troubles me.

And finally, the thing that pushed away with me was the ASSURE data. I have problems consolidating and compromising the data from the ASSURE trial.

Susan Halabi
Professor with tenure, Department of Biostatistics and Bioinformatics, Duke University Medical Center

B. FDA Medical Decision Making

- Note that the trial would not be submitted to FDA if the sponsor/company did not believe there was a decent likelihood of approval.
  - *If p-value were 0.06, how would the company have proceeded?*
- What's the evidence that ODAC used (based on their comments)?
  - Reliance on the trial design, which defined a level of evidence and a primary endpoint.
  - Efficacy (DFS and OS)
  - Toxicity/Safety
  - Level of evidence (based on 257 events in two arms)
  - Choice of endpoint
  - Prior trial results (informally Bayesian)
  - Patient need
- Not used? Financial costs.

ASCO

Different decision making scenarios

A. Do the results of my phase II trial of A vs. B provide sufficient evidence to justify proceeding to the phase III setting?
B. Do the results of this phase III trial provide sufficient evidence for the FDA to approve treatment regimen B?
C. Do the results of this phase III trial provide sufficient evidence for a doctor to prescribe regimen B instead of regimen A to her patient?
D. Do the results of this phase III trial provide sufficient evidence for single payer systems (e.g. Europe) to justify providing regimen B to its patients?
E. Do the results of this phase III trial provide evidence and information regarding who should receive regimen B (vs A?) in resource-poor settings?

ASCO

Medical Decision Making for Physicians

- **ASCO's mission:** To conquer cancer through research, education, and promotion of the highest quality patient care.
- "Toward fulfillment of this goal and at the direction of its board of directors, the ASCO Value in Cancer Care Task Force set out to *develop a framework that would enable a physician and patient to assess the value of a particular cancer treatment regimen given the patient's individual preferences and circumstances."*
- Task force created "Net Health Benefits Scores"

VOLUME 34 · NUMBER 24 · AUGUST 20, 2016

JOURNAL OF CLINICAL ONCOLOGY

ASCO SPECIAL ARTICLE

Updating the American Society of Clinical Oncology Value Framework: Revisions and Reflections in Response to Comments Received

Lowell E. Schnipper, Nancy E. Davidson, Dana S. Wollins, Douglas W. Blayney, Adam P. Dicker, Patricia A. Ganz, J. Russell Hoverman, Robert Langdon, Gary H. Lyman, Neal J. Meropol, Therese Mulvey, Lee Newcomer, Jeffrey Peppercorn, Blase Polite, Derek Raghavan, Gregory Rossi, Leonard Saltz, Deborah Schrag, Thomas J. Smith, Peter P. Yu, Clifford A. Hudis, Julie M. Vose, and Richard L. Schilsky

Calculating NHB scores in advanced disease (an excerpt)

- Observed range ~ 5 to 85.
- Key elements:
  - Efficacy
  - Toxicity
- Bonus points:
  - QoL (p-value!)
  - Disease-free interval
  - Palliation of symptoms
- Final step: Calculate cost.
- Can argue about dimensionality, but a step in the right direction.
- Additional algorithm for adjuvant setting.

Step 1: Determine the regimen's CLINICAL BENEFIT		
1.A. Is hazard ratio (HR) for death reported?	**YES.** Assign an HR Score for death by subtracting the HR from 1, and then multiplying the result by 100. Write this number in the box labeled "HR Score (death)." **Proceed to 1.F.**	**HR Score (death)**
	No. Proceed to 1B.	
1.B. If HR for death is not reported, is median overall survival (OS) reported?	**YES.** Assign an OS Score by calculating the percentage (ie, fractional) difference in median overall survival between the two regimens and multiply the result by 100. Write this number in the box labeled "OS Score." **Proceed to 1.F.**	**OS Score**
	NO. Proceed to 1.C.	
1.C. If OS data are not reported, is hazard ratio (HR) for disease progression reported?	**YES.** Assign an HR Score for disease progression by subtracting the HR from 1, multiplying the result by 100, and then multiplying this number by 0.8. Write this number in the box labeled "HR Score (progression)." **Proceed to 1.F.**	**HR Score (progression)**
	NO. Proceed to 1.D.	
1.D. If HR for disease progression is not reported, is median progression-free survival (PFS) reported?	**YES.** Assign a PFS Score by calculating the percentage (ie, fractional) difference in median progression-free survival between the two regimens and multiply the result by 100. Multiply this number by 0.8. Write this number in the box labeled "PFS Score." **Proceed to 1.F.**	**PFS Score**
	NO. Proceed to 1.E.	
1.E. If median PFS is not reported, is response rate (RR) reported?	**YES.** Assign an RR Score by adding the complete response (CR) and partial response (PR) rates, multiply by 100, then multiply this number by 0.7. Write this number in the box labeled "RR Score." **Proceed to 1.F.**	**RR Score**
1.F. Calculate the Clinical Benefit Score	Insert the score for HR death, HR PFS, median OS, or median PFS. **Note: You should have a score for only 1 of the clinical benefit scales above.** Write the total in the box labeled "Clinical Benefit Score." **Proceed to Step 2.**	**Clinical Benefit Score**
Step 2: Determine the regimen's TOXICITY		
Does the new regimen represent an improvement in	For each of the regimens being assessed, compare the number and frequency of clinically relevant toxicities, and assign a Toxicity Score) as shown below. Each clinically meaningful toxicity (ie, exclude laboratory results only) is assigned a score between 0.5 and 2.0	**Toxicity Score**

Medical Decision Making for Physicians

- At this point, the drug has already been FDA approved.
  - The doctor (and patient?) know that it has been deemed efficacious by some metric.
  - Doctor (and patient?) would infer that a p-value was small (<0.05) compared to some other therapy.
- The question is whether or not the regimen is the right treatment choice for the patient.
- Who is making the decision?
  - The patient, in consultation with the care provider

ASCO

Medical Decision Making for Physicians: Statistical Evidence?

- Statistical Evidence?
  - **NHB: A validated measure of health benefit based on completed late stage trials**
- Other evidence:
  - Toxicity profile, demonstrating risks
  - Efficacy profile, demonstrating personal benefit
    - **Expected increase in survival**
    - Hazard ratios—not so helpful.
- Are these ‘evidence’?
  - Some yes, some no.
  - At the point of patient care, there has been evidence of efficacy
  - It’s the trade-offs that are critical to the decision
- Major trade-off: COST.
  - Not uncommon for a cancer treatment to cost $100,000 per year.

ASCO

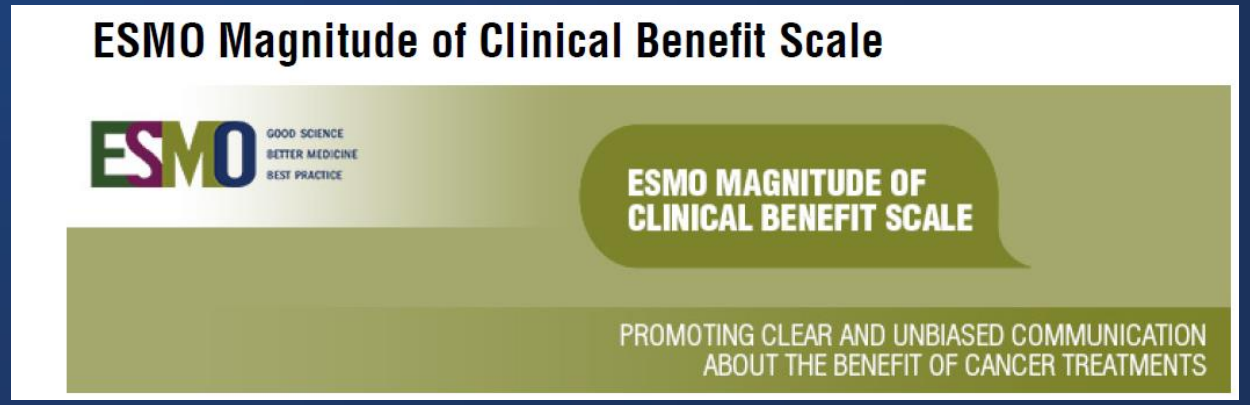
Different decision making scenarios

A. Do the results of my phase II trial of A vs. B provide sufficient evidence to justify proceeding to the phase III setting?
B. Do the results of this phase III trial provide sufficient evidence for the FDA to approve treatment regimen B?
C. Do the results of this phase III trial provide sufficient evidence for a doctor to prescribe regimen B instead of regimen A to her patient?
D. Do the results of this phase III trial provide sufficient evidence for single payer systems (e.g. Europe) to justify providing regimen B to its patients?
E. Do the results of this phase III trial provide evidence and information regarding who should receive regimen B (vs A?) in resource-poor settings?

ASCO

Single Payer System Decision Making (briefly)

- European Society of Medical Oncology (ESMO) has a similar scale to the ASCO NHB scoring system
- Magnitude of Clinical Benefit (MCB) Score
  - Likert scale: 1 to 5
  - Preliminarily there is decent agreement between NHB and MCB scores
- But, there have different questions:
  - Should our single payer system support the use of this agent for patients with this diagnosis?
  - Is the benefit worth the financial cost?
  - What is a fair price for this regimen, given the expected benefit?



Resource-poor settings relying on state care

- Example*: “It is not uncommon to administer trastuzumab only to a certain arbitrary financial point rather than according to standard of care guidelines, which currently suggest 1 year of adjuvant therapy. In such a cost-conscious environment, all attempts should be made to more clearly define a subset of patients who may benefit from costly therapies, thereby improving access by all women regardless of financial or social means.”
- The decision is not to define efficacious treatments, or to help patient make a decision based on trade-offs.
- The decision is to determine how to best ration treatment when treatment is limited.
- Statistical evidence? Depends on the optimization criteria and constraints.
  - Number of lives saved?
  - Fixed vs. variable length of treatment?

*Myburgh et al, (2017) Journal of Global Oncology

ASCO

Role for Statistics in Decision Making in Medical Oncology

- As statisticians, we need to appreciate what decisions need to be made.
  - Efficacy declaration is just a piece of it.
  - Patients, physicians, pharma, payers, regulators care about many more components.
- Huge amount of decision making occurs well-beyond the point of the declaration of efficacy and is based on statistical science.
- Areas of focus for statistics to assist in making those decisions:
  - We need better outcome measures!
    - Development of validated scores (scales) for health benefits
    - Better approaches to measuring tumor burden
  - Predictive modeling for costs and trade-offs
  - Data visualization for ease of interpretation of results

ASCO