Biostatistics in Prostate Cancer Research

Elizabeth Garrett-Mayer, PhD
Associate Professor of Biostatistics and Epidemiology
Director of Biostatistics, Hollings Cancer Center

June 11, 2009
Statistics

- Statistics is the art/science of **summarizing data** and quantifying evidence
- Better yet...summarizing data so that non-statisticians can understand it
- Scientific investigations usually involve collecting a lot of data.
- But, at the end of your study, what you really want is a “punch-line:”
  - Did the new treatment work?
  - Are the two groups being compared the same or different?
  - Is the new method more precise than the old method?
- Statistical inference is the answer!
How do statisticians help research?

- Statistics should be a part of the study from the very beginning
- Statistical issues arise in:
  - Study Design
  - Analysis
  - Interpretation of results
  - Conclusions
What we do

- We plan
- We estimate
- We test
What we do

- **We plan**
  - we help to plan clinical trials and other kinds of studies
  - we help figure out how many people to study

- **We estimate**
  - we determine what the “response rate” was
  - we estimate how much better treatment A is than treatment B

- **We test**
  - we determine which treatment is better
  - we quantify how much better using a test.
Clinical Research in Prostate Cancer

- Research requires a **plan**
- A DETAILED plan called a “clinical trial protocol”
  - could also be an intervention
  - could also be an observational study
  - but, for simplicity, we focus on a “treatment trial”
  - Example: Velcade for treatment of men with relapsed prostate cancer
Clinical Trial Protocol

- Variety of templates
- Some key elements
  - **Specific Aims:** you must state what your goals are in terms of measurable objectives
  - **Background/Rationale:** explanation of why this study is important, what preliminary data exists and justification of the dose.
  - **Experimental Design:** Describes how the study will proceed. no detail can be spared. someone else should be able to implement the study with no questions.
  - **Analysis Plan:** how will the data will handled and objectives answered.
Clinical Drug Trial Checklist.

1 Study Title
2 Study personnel
3 Rationale
4 Objectives
5 Study Plan & Schedule of Assessments
   5.1 Methods of collecting data
   5.2 Study Plan
   5.3 Schedule of Assessments
6 Inclusion Criteria
7 Exclusion Criteria
8 Prohibited Drugs and Interventions.
9 Study design and analysis
   9.1 Randomisation
   9.2 Power calculations
   9.3 Data to be analysed
   9.4 Analysis populations
   9.5 Withdrawals (protocol violations, broken blinding, withdrawal)
   9.6 Statistical Analysis
   9.7 Interim analyses
10 Safety: Reporting of Adverse Events
   10.1 Definition of adverse events provided
   10.2 Investigator's responsibility to report adverse events
   10.3 Definition of serious adverse events in accordance with standard criteria
   10.4 Investigator's responsibility to follow-up and characterise adverse events
   10.5 Procedures for informing CDTC/RIEC of adverse events reports
11 Pharmacy issues: drug storage, dispensing and labelling
12 Administrative issues
13 Compliance With Good Clinical Practice, Ethical Considerations & Informed Consent
Endpoint selection

- What measures should we take to determine if our treatment (e.g. Velcade) has worked?
- Example: for each patient, determine if his disease has
  - regressed?
  - stayed the same? (‘stable disease’)
  - progressed?
- Common endpoints in prostate cancer clinical trials
  - PSA (prostate specific antigen), a biomarker
  - tumor size/volume
  - pain
  - quality of life
- It is important to use endpoints that everyone else uses.
Statistical Design Issues

- Choose most efficient design
- Consider all aims of the study
- Particular designs that might be useful
  - Cross-over
  - Pre-post
  - Factorial
- Sample size considerations
- Interim monitoring plan
Example: prostate cancer clinical trial

- **TAX327**: Aventis study
- **Patient Population**: *hormone refractory metastatic prostate cancer*
- **Large randomized clinical trial**
  - docetaxel, schedule 1
  - docetaxel, schedule 2
  - mitoxantrone
- **Primary endpoint**: overall survival
- **Additional Aim**: how is PSA related to overall survival?
  - prostate specific antigen
  - well-known ‘surrogate’ for prostate cancer presence
  - well-known ‘test’ for prostate cancer progression
- **Additional Aim**: compare quality of life in the three treatment arms
Study design

- Patients are randomized to one of three arms
- Equal chance of assignment to each arm
- Overall survival:
  - Time from randomization until death
  - Patients are followed until death
  - For patients who do not die by study end, we say that their outcomes are ‘censored’ at the last known time they were still alive (more on that later)
- Statistician worked with the clinicians to determine how many patients were needed
  - depends on how certain we want to be about our conclusion
  - the expected survival in each group
  - how long patients are followed
  - how long it takes to enroll patients
Analysis Plan: Part of the Design!

• Statistical method for EACH aim
• Account for type I and type II errors
  ▪ these quantify how certain we want to be about making mistakes
  ▪ type I: the probability of concluding that there is a difference in treatments when there truly is no difference
  ▪ type II: the probability of concluding that there is no difference when there truly is a difference
• Stratifications or adjustments are included if necessary
• Simpler is often better
• Loss to follow-up: plan for missing data
Estimation

- At the end of the study, you need to be able to “measure” how things went

- Some examples:
  - what proportion of patients responded to the treatment?
  - how many patients are still alive at 5 years?
  - what is the difference in the response rate between the two treatment groups?
  - how much improvement was seen in quality of life from the beginning of the study to the end?

- Estimation depends on the endpoint selection
Estimation in TAX 327

- Outcome of interest is overall survival
- We can estimate
  - median survival: the time at which 50% of patients are still alive
  - 5 year survival: the proportion of patients that are still alive at 5 years
- These are called “point estimates”
- Other aims?
  - the mean change in quality of life from baseline to follow-up
  - the proportion of men with increased PSA at end of treatment
Median survival

- Docetaxel every 3 wks: Median survival = 19.4 months

- Docetaxel weekly: Median survival = 18.7 months

- Mitoxantrone: Median survival = 16.6 months

- Which looks to be the best?
Another key part of estimation

- **Precision**: how certain are we of our point estimates?
- Variance or standard errors are important!
- We often use ‘Confidence intervals” to describe our certainty in our estimates
- **A 95% confidence interval**: provides an interval that we are 95% certain contains the true parameter estimate
- 95% is most common, but we also see 90% and 99%.
## Confidence intervals for Median survival in TAX327

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>median</th>
<th>0.95LCL</th>
<th>0.95UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doce Q3</td>
<td>241</td>
<td>19.4</td>
<td>17.6</td>
<td>21.6</td>
</tr>
<tr>
<td>Doce wk</td>
<td>217</td>
<td>18.7</td>
<td>16.3</td>
<td>21.2</td>
</tr>
<tr>
<td>Mitox</td>
<td>228</td>
<td>16.6</td>
<td>14.3</td>
<td>18.6</td>
</tr>
</tbody>
</table>

How to interpret these?
Testing

- Critical for these types of comparative studies!
- The drug company (and everyone else) wants to know if its drug is better than the old drug.
- We test hypotheses:
  - hypothesis 0: survival is the same in the three groups
  - hypothesis 1: survival is different in the three groups.
- Depending on the type of outcome, we use different tests.
- hypothesis 0 is called the “null”
- hypothesis 1 is called the “alternative”
Outcome of test: p-value

- The most common measure of whether or not the treatments are different is the ‘p-value’
- The p-value is the probability of observing the difference we did (or larger) *if the null hypothesis is true.*
- If the p-value is small, it means that the observed data is unlikely if there is really no difference
- If the p-value is large, it means that the observed difference is too small to provide evidence of a “real” difference
- Standard threshold for “significant” p-value?
TAX327

- The ‘logrank test’ is a type of test we use for testing overall survival

- The p-value for testing that all groups are the same is 0.007

- The p-value testing that survival in the Doce Q3 arm is the same as the Doce every week arm is 0.37
- The p-value testing that survival in the Doce Q3 arm is the same as the Mitox arm is 0.009
- The p-value testing that survival in the Doce every week arm is the same as the Mitox arm is 0.10
Additional biostatist issues in prostate cancer research

- Measure of ‘response’
- Measuring time to progression or time to death
Prostate Specific Antigen

- **Prostate specific antigen (PSA)** is a protein produced by the cells of the prostate gland.
- PSA is present in small quantities in the serum of normal men, and is often elevated in the presence of prostate cancer and in other prostate disorders.
- A blood test to measure PSA is considered the most effective test currently available for the early detection of prostate cancer, but this effectiveness has also been questioned.
- Rising levels of PSA over time are associated with both localized and metastatic prostate cancer.
Prostate Specific Antigen (PSA)
Tricky issues with PSA

- Change in PSA from baseline to post-treatment
- Potential problems
  - There is variability due to things other than cancer
    - day to day fluctuations
    - assay sensitivity
    - other prostate disorders
  - When you sample may give you different answers
  - Some question whether or not PSA is a good “surrogate measure”
Surrogate measure

- What is the gold-standard measure in cancer treatment?
- Multiple choice:
  A. time from treatment until disease goes into remission
  B. time from diagnosis until disease progresses
  C. time from treatment until death
  D. time from diagnosis until death
  E. time from treatment until disease progresses
  F. time from diagnosis until disease goes into remission
Surrogate measures in cancer research

- We generally assume the following:
  - if we can shrink the tumor, we can extend life
  - if we can delay tumor progression, we can extend life
- Are these valid assumptions?
  - sometimes yes, sometimes no
- Tumor shrinkage ("clinical response")
  - tumor response is often considered a poor surrogate
- Time to progression
  - tumor progression is often valid surrogate
  - however, it is hard to measure
RECIST criteria

- RECIST criteria offer a simplified, conservative, extraction of imaging data for wide application in clinical trials. They presume that linear measures are an adequate substitute for 2-D methods and registers four response categories:
  - CR (complete response) = disappearance of all target lesions
  - PR (partial response) = 30% decrease in the sum of the longest diameter of target lesions
  - PD (progressive disease) = 20% increase in the sum of the longest diameter of target lesions
  - SD (stable disease) = small changes that do not meet above criteria

http://imaging.cancer.gov/clinicaltrials/imaging/
Potential Problems with RECIST

- Stable disease includes both improvements and worsening
- Tumors are 3-D. RECIST only allows for 1-D. Measures are hence fraught with measurement error.
- Tumors with minor differences (e.g., 32% decrease and 28% decrease) are categorized differently.
Time to event outcomes

- In cancer research, we are usually interested in measuring time until an event occurs.
- The event is usually bad so we are trying to prevent the event from occurring.
- Inevitably, at the end of the study, many patients will not have had the outcome.
- This is called ‘censored’.
- More specifically, “right censored”.
Simple example:
Introduce “administrative” censoring

Time 0

STUDY END
Introduce “administrative” censoring

Time 0

STUDY END
More realistic: clinical trial

Time 0

STUDY END
More realistic: clinical trial

Time 0

STUDY END
Additional issues

- Patient drop-out
- Loss to follow-up
Drop-out or LTFU

Time 0

STUDY END
How do we ‘treat” the data?

Shift everything so each patient time represents time on study.
Set of tools for time-to-event outcomes

- “Survival analysis”
- Kaplan-Meier curves: graphical representation
- Kaplan-Meier estimation: provides point estimates and confidence intervals
- Logrank test: tests for differences across groups
Kaplan-Meier curves

Log Rank p-value = 0.007
Summary

- Biostatisticians have a lot of tools for helping with prostate cancer research
- Critical areas of assistance:
  - study design
  - sample size estimation
  - data analysis
- Prostate cancer has some specific areas that make it challenging
  - measurement issues with standard outcomes
  - time to event outcomes require special methods