Cancer Biostatistics Workshop
Science of Doing Science - Biostatistics

Yu Shyr, PhD

Jan. 18, 2008

Cancer Biostatistics Center
Vanderbilt-Ingram Cancer Center

Yu.Shyr@vanderbilt.edu
Aims

- Cancer Biostatistics Workshop Introduction.

- Issues of the experiment design of biomedical research – adaptive, Bayesian.

- P-value, type I error, FDR and methods of multiple comparison.

- Topics of the high-dimensional data experiment including the assessment of the quality control.

- Tools of the visualization.
Series One:

What You Probably Don’t Know About Biostatistics

January 18th - Yu Shyr, PhD
“The Science of Doing Science – Biostatistics”

February 15th – Dan Ayers, MS
“Skittles, An Ounce of Measurement, and Intentional Science”

March 21st – Ayumi Shintani, MPH, PhD
“Introduction to Power and Sample Size Estimation”
Series Two:

Essentials of Efficient Experimental Design

April 18th - Tatsuki Koyama, PhD
“Phase II Clinical Trials”

May 16th - Leena Choi, PhD
“Adaptive Designs in Dose-Finding (Phase I) Studies: A Bayesian Approach”

June 20th - Fei Ye, PhD
“Phase III Clinical Trials”
Series Three:

High-Dimensional Data

August 15th - Chun Li, PhD
“Genome-wide Association Analysis for the Shanghai Breast Cancer Study”

September 19th - Irene Feurer, PhD
“An Introduction to Principal Component Analysis and Data Reduction Methodologies”

October 17th - Ming Li, PhD
“Statistical Analysis Strategies for MALDI-TOF and Shotgun Proteomics Data”
Special Topics:
Observational Data and Epidemiology

November 21st - Ayumi Shintani, MPH, PhD
“Introduction to Observational Study Design”

Please call 6-2502 or e-mail shaun.m.haskins@vanderbilt.edu
if you have any questions or feedback.
Type of the Biomedical Studies

- **in vitro** – cell lines
- **in vivo** – animal models
- **Clinical Trials** – Phase I, II, III, IV
- **Cohort Study** – Prospective, Retrospective
- **Case-Control Study** – Nested Case-Control
Experiment Design

- Primary objective of the study
- Primary hypothesis of the study
- Primary endpoint of the study
- How many groups? How many time points?
- Preliminary data
Adaptive trial design refers to a clinical trial methodology that allows trial design modifications to be made after patients have been enrolled in a study, without compromising the scientific method. In order to maintain the integrity of the trial, these modifications should be clearly defined in the protocol. When designed well, an adaptive trial empowers sponsors to respond to data collected during the trial. This is achieved by re-focusing the trial in a way that maximizes the impact of each subject’s contribution.

Examples of adaptive trial designs include dropping a treatment arm, modifying the sample size, balancing treatment assignments using adaptive randomization or simply stopping a study early for success or failure.
Two Stage Design

One of the primary objectives of a phase II clinical trial of a new drug or regimen is to determine whether it has sufficient biological activity against the disease under study to warrant more extensive development. Several two-stage Phase II designs will be discussed here –

Gehan,

Fleming,

Simon’s Optimal,

Simons’s Minimax,

Balanced Design
Gehan’s (1961) design is the oldest design. It is a two-stage design for estimating the response rate but providing for early termination if the drug shows insufficient antitumor activity. The design is most commonly used with a first stage of 14 patients. If no responses (completed response or partial response) are observed, the trial is terminated. The rationale for stopping is that if the true response probability were at least 20%, at least one response would very likely have been observed in the first 14 patients. If no responses are seen, it is unlikely that the true response probability is at least 20%.
If at least one response is observed in the first 14 patients, then a second stage of accrual is carried out to obtain an estimate of the response rate. The number of patients to accrue in the second stage depends upon the number of responses observed in the first stage and upon the precision desired for the final estimate of response rate. If the first stage consists of 14 patients, the second stage will consist of between one and 11 patients if a standard error of 10% is desired. The second stage will consist of between 45 and 86 patients if a standard error of 5% is desired.
Simon’s Optimal Design

The optimal two-stage designs (Simon, 1989) are optimal in the sense that the expected sample size is minimized if the regimen has low activity subject to constraints upon the size of the type I ($\alpha$) and type II ($\beta$) errors.

Example of Simon’s Optimal Design:
Clinical uninteresting level = 5% response rate,
Clinical interesting level = 20% response rate,
type I error ($\alpha$) = 0.05
type II error ($\beta$) = 0.20 or Power = 0.80 (Power = 1 – type II error).

Stage I Reject drug if response rate $\leq 0/10$
Stage II Reject drug if response rate $\leq 3/29$
Frequentist versus Bayesians

Frequentists: $\alpha$ and $\beta$ errors

Bayesians:
- Quantify designs with other properties
- General philosophy
  - Start with prior information (“prior distribution”)
  - Observe data (“likelihood function”)
  - Combine prior and data to get “posterior” distribution
  - Make inferences based on posterior
Bayesian inference

- No p-values and confidence intervals
- From the posterior distribution:
  - Posterior probabilities
  - Prediction intervals
  - Credible intervals
- Bayesian designs
  - Can look at data as often as you like (!)
  - Use information as it accumulates
  - Make “what if?” calculations
  - Helps decide whether to stop now or not
Bayesian Designs

- Requires ‘prior’
  - Reflects uncertainty about the response rate
  - Can be ‘vague’, ‘uninformative’
  - Can be controversial: inference may change
Bayesian Designs
Posterior Probabilities

- Blue line: probability $p < 0.20$
- Red line: probability $p > 0.40$

Cumulative Number of Patients vs. Probability
Other priors

- What if we had used a different prior?
- Assume informative “orange” prior
Variability of *In Situ* Proteomic Profiling and Implications for Study Design in Colorectal Tumors

Overview

- The comprehension of intrinsic **tumor heterogeneity** is vital for understanding of tumor progression mechanisms as well as for providing efficient treatments.

- *In situ* proteomic profiling of tumors is a powerful technology with potential to enhance our understanding of tumor biology, but **sources of variability due to patient and tumor heterogeneity** are poorly understood and are the topic of this investigation.
Questions

• Is proteomic profiling in adenomas more homogenous across patients than in normal mucosa specimens?

• Does primary carcinoma exhibit greater heterogeneity than normal mucosa and adenomas?

• Is Inter- and intra-case variability approximately equal for protein peaks?

• What is the optimal number of sub-samples per case that will reduce the total number of cases required?

• How to characterize intra- and inter-case variability of high-throughput protein expression in colorectal tumors?
Table IV
Power = 80% Type I error = 5%

<table>
<thead>
<tr>
<th>Subsample (m)</th>
<th>Number</th>
<th>Inter-Case Variance</th>
<th>Inter-Case Variance</th>
<th>Inter-Case Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.2</td>
<td>0.5</td>
<td>1.0</td>
<td>0.2</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>11</td>
<td>19</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>9</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>4</td>
</tr>
</tbody>
</table>
P-value, Type I error, FDR, and multiple comparison

**Type I error:** For testing a null hypothesis, $H_0$, against an alternative hypothesis, the type I error of the test is the rejection of the null hypothesis when it is true.

The *p value* of a statistical test is the probability of observing a sample outcome as extreme or more extreme as the one observed. This probability is computed under the assumption that the null hypothesis is true.
Linear Regression

Graphs by group
### Linear Regression

**Regression Results:** Slope $0.5 (p=0.0022)$

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob > |T| |
|----------|----|-------------------|----------------|-----------------------|--------|---|
| INTERCEP | 1  | 3.000091          | 1.12474681     | 2.667                 | 0.0257 |
| X        | 1  | 0.500091          | 0.11790550     | 4.241                 | 0.0022 |

**Which graph on the previous slide corresponds to this analysis?**
Solutions for the Multiple Comparison Problem

- A MCP Solution Must Control False Positives
  - How to measure multiple false positives?

- Familywise Error Rate (FWER)
  - Chance of *any* false positives
  - Controlled by Bonferroni & Random Field Methods

- False Discovery Rate (FDR)
  - Proportion of false positives *among* rejected tests
False Discovery Rate

- **Observed FDR**
- **Obs FDR** = \( \frac{V_{0R}}{V_{1R} + V_{0R}} = \frac{V_{0R}}{N_R} \)
- **If** \( N_R = 0 \), \( \text{obsFDR} = 0 \)
- **Only know** \( N_R \), **not how many are true or false**
- **Control is on the expected FDR**
- **FDR** = \( E(\text{obsFDR}) \)
Benjamini & Hochberg Procedure

- Select desired limit $q$ on FDR
- Order $p$-values, $p_{(1)} \leq p_{(2)} \leq ... \leq p_{(v)}$
- Let $r$ be largest $i$ such that
  \[ p_{(i)} \leq \frac{i}{v} \times \frac{q}{c(v)} \]
- Reject all hypotheses corresponding to $p_{(1)}, \ldots, p_{(r)}$.
Benjamini & Hochberg: Key Properties

- FDR is controlled
  \[ E(\text{obsFDR}) \leq q \frac{m_0}{\text{V}} \]
  - Conservative, if large fraction of nulls false

- Adaptive
  - Threshold depends on amount of feature
    - More feature, More small \( p \)-values,
      More \( p_{(i)} \) less than \( \frac{i}{\text{V}} \times \frac{q}{c(V)} \)
Benjamini & Hochberg Procedure

- \( c(V) = 1 \)
  - Positive Regression Dependency on Subsets
  - \( P(X_1 \geq c_1, X_2 \geq c_2, ..., X_k \geq c_k \mid X_i = x_i) \) is non-decreasing in \( x_i \)
  - Only required of test statistics for which null true
  - Special cases include
    - Independence
    - Multivariate Normal with all positive correlations
    - Same, but studentized with common std. err.

- \( c(V) = \sum_{i=1}^{\sqrt{V}} \frac{1}{i} \approx \log(V) + 0.5772 \)
  - Arbitrary covariance structure

*Ann. Stat.*
29:1165-1188
The major challenge in high throughput experiments, e.g., microarray data, MALDI-TOF data, Array CGH data, Shotgun Proteomic data, or SNP data, is that the data is often high dimensional.

When the number of dimensions reaches thousands or more, the computational time for the pattern recognition algorithms can become unreasonable. This can be a problem, especially when some of the features are not discriminatory.
Contributions to uncertainty can come from a variety of resources:

- Analytic variation
- Sampling methods
- Sample preparation
- Biological Heterogeneity in Population
- Specimen Collection/Handling Effects
  - Tumor: surgical related effects
  - Cell Line: culture condition
- Biological Heterogeneity in Specimen
ISO 17025 5.4.6.3 “When estimating the uncertainty of measurement, all uncertainty components which are of importance in the given situation shall be taken into account using appropriate methods of analysis.”

- repeatability standard deviation (intra-lab)
- reproducibility standard deviation (inter-lab)
- intermediate precision (more factors in the system)
The main objectives in performing an R&R study are to identify and quantify the absolute and relative contribution of each source of variation, to decide if the measurement process is adequate or not and, if not, to correct the errors by recalibrating the instrument, training the operators, other mathematical corrections, etc.

In quality control and quality assurance, the goal is to reduce the variability in the system and, consequently, the variability in the biomarker measurement.
Quality Control Assessment
Reproducibility

- Correlation of Variation (CV)
  SD/Mean

- Intra-class Correlation Coefficient (ICC)
  Intra / Intra + Inter

- Variance Component Analysis
  Mixed/Random Effect Model.
  The model: investigators, day, spot, machine, lab, etc.

- Goal – Make sure the data is reproducible !!
WFCCM – Class Prediction Model

Accuracy vs Features

Training data set - 93 vs 91

Testing data set - 52 vs 58
Receiver Operating Characteristic (ROC) Curve

Training

Test

AUC=0.822

AUC=0.819
WFCCM – Class Prediction Model
Receiver Operating Characteristic (ROC) Curve

Training

AUC=0.822

Test

AUC=0.654
Hierarchical clustering of gene expression data
Multidimensional scaling (MDS)
Multidimensional Scaling
Multidimensional Scaling

Response (A + C):
- PR
- SD
- PD

Survival (B):
- SD, OS > median
- SD, OS < median
- PR, OS > median
- PD, OS < median
Dear Dr. Shyr,

I am a fellow in the Department of... We put together a paper recently ... I had performed the statistical analysis incorrectly... A reviewer pointed out that with so many t tests some might give a ‘significant’ result by chance and suggested we present ANOVA F statistics instead...we will pay the full cost of any such work you may perform...

Thanks and Regards
Dr....
When Do You Need a Biostatistician?

- Study Design
  - Implementation
  - Statistical Data Analysis
    - Abstract Writing
    - Paper Writing
    - Letter to the paper reviewer
END