Statistical Analyses in Medicine.

RICHARD P. WATERMAN
Department of Statistics
The Wharton School
University of Pennsylvania

14th November 1995

Introduction

Aims:

- Discuss the use of statistics in medicine.
  - Empirical research paradigm.
  - Where the dialogue between M.D. and Statistician is most effective.
- Update on newer modalities in statistics.
  1. Missing data.
  2. Repeated categorical data, longitudinal studies.
  3. Exact tests, small sample statistics.
  4. Data mining, knowledge seeking from large data bases. Neural nets, CART (Classification and regression trees)
  5. Censored data – Survival analysis, Aids
  6. Meta analysis

Motivation

1. To read papers, critique methodology and understand results.
2. To undertake quality research. (Reproducible)
3. To get funding. (NIH proposals, power analyses). e.g. Penn cancer group.
4. Understanding statistics can lead to more efficient, less expensive and more informative studies.

The roles of statistics

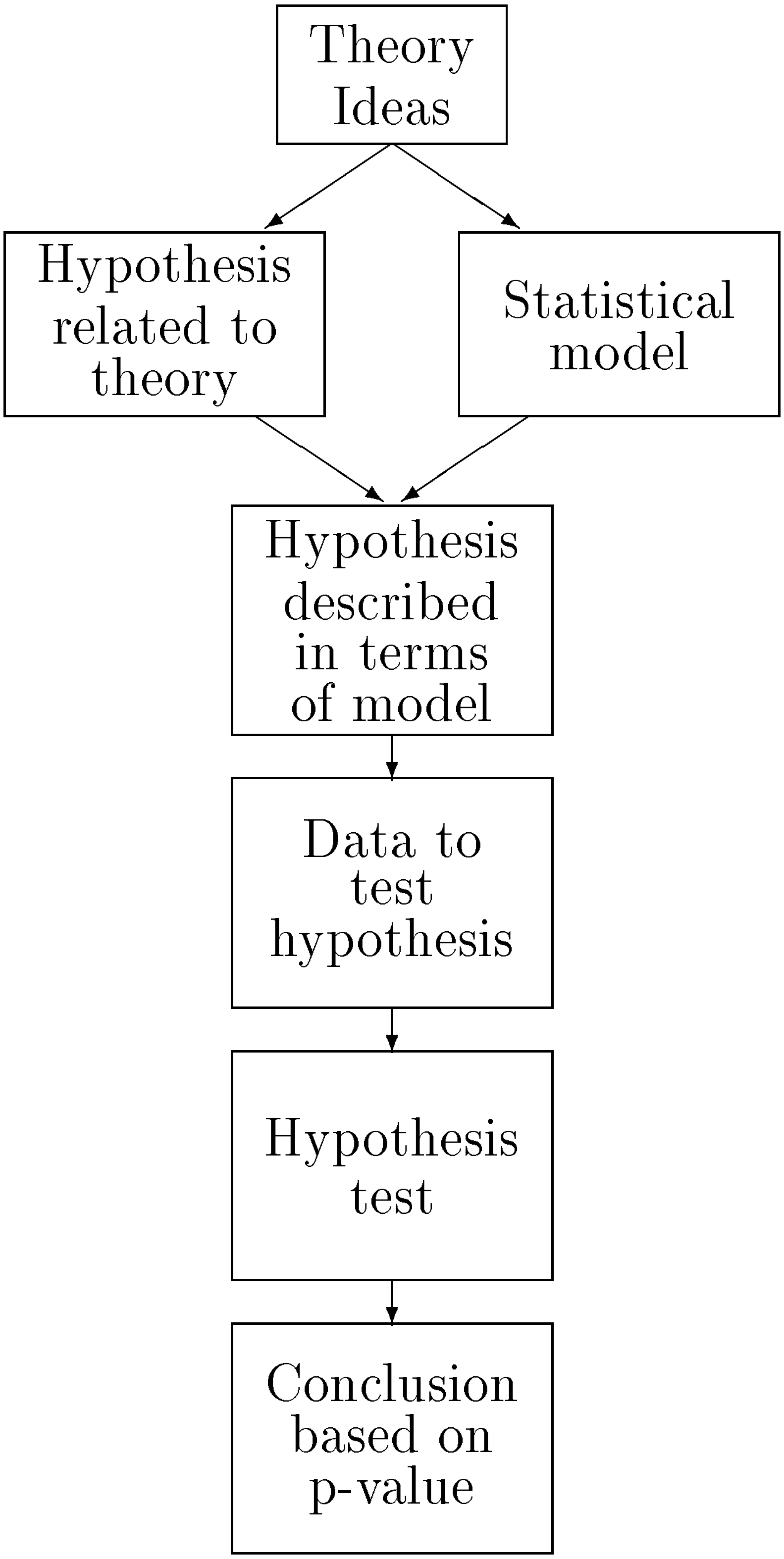
- Summarizing data **without** losing information (descriptive statistics: plots; means *etc*).
- Making decisions.
  - Diagnosis.
  - Research (hypothesis testing).
  - Process control, quality control, monitoring.

Focus today on statistics in **research**

Peppered with **p-values**, the currency of empirical science. A measure of evidence against the null hypothesis.

- Drive our decisions and conclusions.
- Across most applied empirical research.
- What do they mean?
- Where do they come from?
- What are their limitations?

The empirical research paradigm



Adding value

- Completely unrealistic to expect to have a grasp on the minutiae of the process (the bottom end).
- We have basically stopped teaching it.
- Just principles.
- A vital component;
  Get the statisticians involved at the beginning. Save headaches later on. Design a simple effective study.
- The **modeling interface** is where great benefit can happen.

Example

Laparoscopic procedures have revolutionized the practice of surgery.

Laparoscopic versus open inguinal herniorrhaphy: Preliminary results of a randomized controlled trial. Barkun, J.S. *et.al.* Surgery 1995 pp. 703–710.

Theory: Less invasive.

Hypothesis related to theory: Quality of life is better. Convalescence time is faster.

Statistical model. Cox's proportional hazard model:

$$P(\text{Finish convalescence in } t+\Delta t \,|\, \text{Not finished at time } t) = \lambda_0(t)\exp\{\beta z(t)\}.$$

ASSUMPTIONS ... (what are they?)

Hypothesis:

$$H_0 : \beta = 0 \qquad vs. \qquad H_1 : \beta \neq 0.$$

Comments

- Critical parts are at the top of the flowchart.
- If this part has gone wrong then the rest is worthless.
- Mistakes here give rise to Order 1 biases $O(1)$. They don't go away as the sample gets larger.
- The calculation of the p-value is based on the statistical model, **different** models mean different p-values.
- Much of statistical part (at bottom) if done badly, results in mistakes that are $O(1/n)$. Often not a big deal.

O(1) Biases

1. **Experimental**: how data was collected, study design etc.
2. **Modeling**: how the data was understood.

Experimental issues

Objective: minimize biases that would invalidate results or what makes for a good study?

1. Presence of a **control group** for comparison.
2. If treatment and control groups are similar then differences in outcome may be **attributed** to the treatment. (`causality`)
3. If control, is **different** from treatment with respect to other factors then their effects are **confounded** with treatment.
4. **Randomizing** individuals to treatment and control ensures no confounding between groups.
5. **Placebo** for the control group.
6. **Double blind** trials (maybe not in surgery!)

The best intentions ...

Randomized controls versus historical controls.

Therapy	*Randomized controlled*		*Historical controls*	
	+	−	+	−
Coronary bypass surgery	1	7	16	5
Anticoagulants	1	9	5	1
5-FU	0	5	2	0
BCG	2	2	4	0
DES	0	3	5	0

Overall 13% of randomized controlled studies found positive results but 84% of historical controls.

Source: Sacks, Chalmers, and Smith. *American Journal of Medicine* Vol. 72 (1982) pp. 233–40

The modeling issues

Objective: relate an outcome variable to explanatory variables (covariates).

- **Outcome** variable(s). Your choice. Survival time or quality of life?
- **Covariates**. Identify factors influencing the outcome. Account for the sources of variability in the outcome. Identification of likely prognostic factors.
- **Functional relationship** between covariates and outcome (is there interaction between the covariates?).
- **Discrepancy** between what the model says should happen and what actually happens. Compare what you observe with what you expect. (model **diagnostics**, goodness of **fit**, **residuals**.)

If you see a p-value then there was a model. How well did it fit?

Example

Laparoscopic versus open inguinal herniorrhaphy: Preliminary results of a randomized controlled trial. Barkun, J.S. *et.al.* Surgery 1995 pp. 703–710.

Does age also have an effect on convalescence time?

Rather than looking at the **marginal** effect of laparoscopy on convalescence time, look at the **partial** effect having included (modeled) the effect of **age**. Univariate vs multivariate models.

Functional form of association (typically linear, additive effects):

$$\text{Impact on convalescence} = \beta_0 + \beta_1 \text{Treatment} + \beta_2 \text{Age}$$

Hypothesis:

$$H_0 : \beta_1 = 0 \qquad vs. \qquad H_1 : \beta_1 \neq 0.$$

Or .. could the effect of Treatment depend on Age?
Interaction.

$$\text{Impact} = \text{Treatment} + \text{Age} + \text{Treatment*Age}$$

Or ... could the effect be "jump like"?

Modeling: good practice

1. The **outcome** variable. Is the ideal **outcome** really uni-dimensional? Does the outcome variable genuinely measure what you are interested in? Consider short term vs. long term effects.

2. The **covariates**. Those most likely related to the outcome plus variables that can be used for comparison: the usual demographics plus measures that can be validated across the literature, (recognize biases in the sample).

3. Think about functional relationships, e.g. smooth or sudden? Could there be **interaction?**

4. Always question assumptions.

5. Check to see if the model fits.

Practical considerations

- Keeping high quality data records. Database management.
- High quality people to look after this end of things. Not a place to save money.

Some points on p-values

Assume a good design and a good model.

Recognize:

1. as the sample size gets big the p-value always becomes significant.
2. If there are many p-values then you tend to see significant things just by chance alone.

E.g. in the laparoscopy study 12 baseline variables were compared. Only one was significant ($p = 0.014$), "Anticipated return to full activities". Discussants pick up on this.

But with twelve comparisons there is a 1 in 6 chance of seeing a p-value this small even if no real difference. So not so surprising.

Practical significance vs. statistical significance.

Summary

- Biases:
  - Design based.
  - Model based.
- The interface where your input is most valuable.
- Recognize the importance of assumptions.
- Make sure someone has checked them.
- Does the model fit?
- Appreciation of p-values.
- Practical importance vs. statistical significance

Current trends in applied statistics – 1

Missing data.

- Not uncommon to lose 30–50% of original sample due to missing data. Outcome variable or some covariates may be missing.
- Problem: ignoring these patients may cause bias. If they are missing for a reason, e.g. non-compliance.
- Problem: waste of resources.
- Approach: understand and **model** the "missingness" mechanism.
- Progress made in the last few years with this.
- Little, Roderick J. A.; Rubin, d,. *Statistical analysis with missing data.*

Current trends in applied statistics – 2

Repeated categorical data.

- Many studies involve “follow up”, so collect repeat data on the same individual.
- Multiple responses recorded on the same experimental unit. e.g. the **unit** is a mouth and the response is measured on teeth.
- This “**clustering**” phenomena usually violates a core statistical assumption of **independence**.
- Much current research (Liang and Zeger, Generalized Estimating Equations) is in this area.

Current trends in applied statistics – 3

Exact tests.

- Studies with small or highly unbalanced samples.
- **Problem** Most statistics assumes large sample sizes (asymptopia) and reasonable balance.
- Uncritical application of standard methods can be very misleading. (p-values all wrong)
- Exact inference is more appropriate. e.g. Fisher's Exact Test for 2x2 tables.
- Now many generalizations to complex statistical models, e.g. Exact conditional logistic regression, `LogXact`, (Mehta et.al.)

Current trends in applied statistics – 4

Data mining.

- Gigabytes of data is now available.
- Medical informatics.
- Trawl through a database and see what's there, what can we learn? No *a priori* hypotheses. A new scientific paradigm. Popular in marketing, finance
- The First International Conference on Knowledge Discovery and Data Mining (KDD-95), held in Montreal in August 1995.
- Methods: Neural networks, Classification and regression trees.

Current trends in applied statistics – 5

Censored data.

- Censoring: only partial information is available. Typically don't know when an event of interest (infection) began/ ended or both!
- "Lost to follow-up" and "End of trail" are the classic examples.
- AIDS related funding has led to lots of new research; many different types of censoring now considered.
- 93Biomtrcs 49 13- 22 J Kim, Mimi Y.;De Gruttola, Victor G.;Lagakos, Stephen W.
  Analyzing doubly censored data with covariates, with application to AIDS

Current trends in applied statistics – 6

Meta analysis.

- If a treatment effect is real but small, or data is very noisy then the treatment effect may go unnoticed.
- Result is a non-significant p-value.
- Putting lots of small studies together can allow significant treatment effects to be identified.
- Technique is called “meta-analysis”
- Analyzing the analyses.
- Requires special statistical procedures.

Summary

- Many new application areas.
- New techniques and models opened up by computational power.