Core Clinical Research Concepts and Issues

AI M Best, PhD
Virginia Commonwealth University

Task Force on Design and Analysis in Oral Health Research
Satellite Symposium, AADR

Boston MA: March 9, 2015

Outline

- Core Concepts
  - Defining the research question
  - Causation vs. association
  - Sources of error in clinical research
  - Multiple outcomes: Primary & secondary outcomes

- Prominent Design Issues in Clinical Research
  - Randomization
  - Masking/blinding
  - Analysis: Intention-to-treat or Per Protocol?
  - Biomarkers and surrogate measures
  - Superiority, equivalence, or non-inferiority

Intro to Multivariate Statistical Modeling

- Regression
  - Continuous outcomes: Linear ...
  - Dichotomous outcomes: Logistic ...
  - Counts: Poisson ...
  - Time to event: Proportional hazards survival
- Unified management of confounding and effect modification in statistical models
- Propensity methods

Core Concepts

Part 1
Defining the research question

- Conceptual progression from general to specific
- General question
  - Can GCF constituent levels distinguish active chronic adult perio sites from inactive and simple gingivitis sites?
- Specific hypothesis
  - Site GCF AST levels with active perio exceed those with inactive perio and simple gingivitis.
- Testable consequence
  - ...

Defining the research question: Testable consequence

- Context
  - precise population: eligibility criteria*
  - sampling: patient sources, recruitment, consent
  - measurement definitions: detailed operational definitions
    - distinguishing active sites, inactive sites, gingivitis sites
    - GCF sampling, storage, and AST measurement procedures
- Prediction: a relationship between exposure and outcome
- Formalization: refutable hypotheses
  - * AKA: inclusion/exclusion criteria

Defining the research question: Prediction

- Prediction: a statistical relationship between exposure and outcome
  - GCF AST will be tend to be higher in active than in inactive or simple gingivitis sites within the same or different patients
- Formalization: a refutable hypothesis

Defining the research question: Formalization

- A refutable hypothesis
- Statistical formalization:
  - Ho: Mean(Active) = Mean(Inactive) = Mean(Gingivitis)
    - Which may be disproved beyond a reasonable doubt through falsification by data via statistical hypothesis testing, in favor of:
  - Ha: Mean(Active) > Mean(Inactive) and Mean(Active) > Mean(Gingivitis)
Defining the research question: Formalization

- Estimating clinically relevant quantities
- Clinical formalization:
  - Mean(Active) – Mean(Inactive) > clinically important \( \delta \) and
  - Mean(Active) – Mean(Gingivitis) > clinically important \( \delta \)
  - Or, ROC AUC for distinguishing active perio > 0.75
- These quantities may be estimated, and the clinically relevant hypotheses proven or disproved, using confidence intervals and associated statistical tests

Characteristics of a well-framed clinical research question

- FINER
  - Feasible
  - Interesting
  - Novel
  - Ethical
  - Relevant

Source: Hulley, et al. (2007) Designing Clinical Research, Lippincot, Table 2.1

... a well-framed clinical research question

- Feasible
  - Adequate subjects
  - Adequate technical expertise
  - Affordable in time and money
  - Manageable in scope: logistics
- Interesting
  - Generalizable to more than 1 application
  - An answer would be intriguing to the investigator and at least some other folks
  - Potentially fruitful, suggesting other hypotheses

... a well-framed clinical research question

- Novel
  - New information
  - Or in need of replication
- Ethical
  - Avoids unacceptable risks
- Relevant
  - How might the various outcomes advance knowledge? Influence practice? Establish policy?
  - Or guide future research?
Causation vs. Association

- What do we mean by "Exposure A causes Disease B"?
- One approach: Logical consideration of evidence favoring causality
  - Consistent association across studies and designs
  - Strong association (big differences, strong trends)
  - Dose response relationship
  - Biologic plausibility

*Note: investigators can come up with a*

Empirical approach:

Statistical association
- Tendency of characteristics to vary together, so information on one allows guessing or predicting the other more accurately
- BUT: there are lots of ways we may be fooled

Associations are common, and
How we measure statistical associations

- Associations are what we observe, as
  - Differences or ratios of:
    - means or medians
    - proportions
    - odds
    - rates
  - Slopes of trends in statistical models
  - Correlation, regression coefficients

- Causation → association, but not the other way

- No measure of association, in itself, implies causation

Example of causal relationships

- We think:
  - influenza deaths less common in immunized groups
  - MI less common in hyperlipidemics on statins than in untreated hyperlipidemics
  - Stroke probability ↑ with ↑ SBP, ↑ LDL
  - ↑ exercise, ↓ DBP
  - ↑ sugar consumption, ↑ DMFS
  - ↑ cigarette smoking, ↑ alveolar bone loss

Example of non-causal or unproven relationships

- We are not sure:
  - Antioxidant supplementation can lower US cancer risk
  - Calcium supplements reduce demineralization after 65
  - Treating chronic periodontitis reduces CVD risk
  - Scaling and root debridement in pregnant women will reduce preterm births
Clinical research is a search for causes and predictors

- Clinical research is about finding and distinguishing causal (or consistently prognostic) associations from those due to
  - Reverse causation
  - Chance
  - Measurement bias
  - Selection bias
  - Confounding

Sources of error in clinical research

- Chance
- Causal A → B
- Statistical Association
- Measurement bias
- Selection bias
- Confounding bias
Reverse causation?
Does the outcome proceed from the clinical variable?

- We observe: Prevalence of a disease is higher in those from whom a microbe is successfully cultured than in those from whom it is not
- Is the microbe statistically associated with the disease
  - because the microbe causes the disease?
  - or because the disease makes the site more hospitable to colonization or infection by the microbe?

Measurement (information) bias

- Definition, Bias:
  – Systematic distortion of the estimated intervention effect away from the “truth,” caused by inadequacies in the design, conduct, or analysis of a trial
- Definition, Measurement bias:
  – Systematic or non-uniform failure of a measurement process to accurately represent the measurement target
  – Examples ...

Measurement (information) bias

- Examples, Measurement bias:
  – different approaches to questioning, when determining past exposures in a case-control study
  – more complete medical history and physical examination of subjects who have been exposed to an agent suspected of causing a disease than of those who have not been exposed to the agent

Measurement bias?

- NHANES III: “We estimate that at least 35% of the dentate US adults aged 30 to 90 have periodontitis, with 21.8% having a mild form and 12.6% having a moderate form”¹
  – Mesial and buccal surfaces
  – Two randomly selected quadrants
  – CAL ≥ 3mm
- Or: Full mouth prevalence = 64.7% vs partial = 39.1%²

Recall bias


- Conclusion: “overall, the benefits of frequent brushing of teeth did not outweigh the damaging effect of frequent sugar consumption”

- Conclusion: “for children who brushed their teeth twice a day or more, consumption of sugars and sugary foods did not appear to be associated with caries”

Harris, et al. (2004) Risk factors for dental caries in young children: a systematic review, Community Dental Health, 21(S); 71-85

Lead-time bias

- Bias in evaluating screening programs
  – If time to outcome (e.g., death) is measured from point of diagnosis, then early diagnosis will increase the estimated survival time by the interval between diagnosis-by-screening and when diagnosis would have occurred by ordinary means
  – This can make early diagnosis appear to increase survival time, even when it has no effect ... or even a damaging effect

Selection bias

- Definition: Bias from the use of a non-representative group as the basis of generalization to a broader population
- Can distort: disease frequency measures, exposure-disease associations
- Example: Estimate prognosis from patients newly diagnosed and infer to patients hospitalized with the disease
  – Newly diagnosed patients have a much narrower spectrum of outcomes

“Selection bias” is also used to describe the systematic error that can occur when allocating patients to intervention groups.
Selection (=sampling) bias

- Definition: When admission rates vary between exposed cases and unexposed controls

Adapted from: Roberts, Spitzer, Delmore, Sackett (1978)
*J Chron Dis* (31) 119-128

Admission rate bias

- AKA: Berkson’s bias

<table>
<thead>
<tr>
<th>Diseases of bones and organs of movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory disease</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Totals</td>
</tr>
<tr>
<td>Relative risk=</td>
</tr>
<tr>
<td>Odds ratio=</td>
</tr>
</tbody>
</table>

Adapted from: Roberts, Spitzer, Delmore, Sackett (1978)
*J Chron Dis* (31) 119-128

Length-time bias

- Bias in evaluating screening programs: two cases
  - a slowing progressing disease vs.
  - a rapidly progressing disease

- In standard clinical practice: The observed proportion of those with the disease will be higher for the rapidly progressing disease because they have a shorter asymptomatic phase

- In a screening program: The observed proportion of those with the disease will be higher for the slowly progressing disease because they are detectable over a wider range of time
Length-time bias

- Patients diagnosed by screening will, as a group, progress more slowly than those diagnosed by conventional means. Even if early treatment has no impact.

Confounding

- Informal definition: distortion of the true biologic relation between an exposure and a disease outcome of interest
- Usually due to a research design and analysis that fail to properly account for additional variables associated with both
  - Such variables are referred to as confounders or, less formally, as lurking variables

Confounding examples

- “Honey, it’s not my fault! It’s the one-night stand gene”
- “The D4 mutation of the human dopamine receptor gene is linked with: gambling, alcoholism, drug use, overeating, political liberalism, passion for horror films, ADHD, extreme extraversion, impulsiveness”

Confounding examples

- Misidentified carcinogen
  - Prior to the discovery of HPV, HSV-2 was associated with the cervical cancer
  - It is now well established that HPV is central to the pathogenesis of invasive cervical cancer. And HSV-2 appears to increase the risk


Perio and CVD

- Cigarette smoking is associated with adult perio and CVD
- This produces an association between perio and CVD
- Control for smoking to see the perio-CVD relationship clearly


Confounding

<table>
<thead>
<tr>
<th>Amount of dental care</th>
<th>One year loss of any teeth</th>
<th>Annual %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Survival</td>
<td>Yes</td>
</tr>
<tr>
<td>Less</td>
<td>20</td>
<td>373</td>
</tr>
<tr>
<td>More</td>
<td>6</td>
<td>316</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>689</td>
</tr>
</tbody>
</table>

Relative risk = 2.73
Odds ratio = 2.82

Note: artificial data
Confounding: Simpson’s paradox

- Statistical associations in entire groups can differ from statistical associations in their subgroups
- Worse: statistical associations in whole groups can be in the opposite direction from statistical associations in every subgroup!!!
Multiplicity effects

N=47 perio, N=20 healthy

Vir.gin.i.a   C o m m o n w e a l t h   U n i v e r s i t y

Comparisons of subgingival microbial profiles of refractory periodontitis, severe periodontitis, and periodontal health using the human oral microbiome identification microarray.

Abstract

BACKGROUND: This study compared the subgingival microbiota of subjects with refractory periodontitis (RPs) vs. periodontal health (PH) using the Human Oral Microbiome Identification Microarray (HOMM).

METHODS: Actinobacillus subgingival plaque samples were taken from 47 subjects with periodontitis and 20 individuals with PH and analyzed for the presence of 300 species by HOMM. The subjects with periodontitis were classified as having RPs (n = 17) based on mean attachment loss (AL) and/or more than three sites with AL > 3 mm after scaling and root planing, access, and confirmatory subgingival plaque sampling and determination of RPs.

RESULTS: More species were detected in patients with disease (GR) than in those with PH by a significant higher frequency of pathogens.

Conclusions: The proliferate of possible comparisons in a trial. Common sources of multiplicity are:

- multiple outcome measures,
- assessment at several time points,
- subgroup analyses, or
- multiple intervention groups

Multiple comparisons: Performance of multiple analyses on the same data. Multiple statistical comparisons increase the probability of a type I error: “finding” an association when there is none

Multiple outcome examples

- Caries
  - visual exam
  - x-ray interpretation
  - Fiberoptic Transillumination (FOTI)
  - Electrical Caries Meter (ECM)
  - DiagnoDent (DD)

- Perio
  - alveolar bone loss
  - clinical attachment level
  - pocket depth
Multiple comparisons

- If we control the chance of a false positive answer in evaluating a single outcome by a conventional statistical method, which permits a 5% chance of claiming a statistical association when none exists,
  - then by the time we’ve compared treatments on 4 outcomes our chance of an error is MUCH HIGHER
- We can’t stop asking questions, so three strategies are used to control this
  - Require stronger evidence to claim an effect

Require stronger evidence to claim an effect

- Example: When asking 5 questions
- Require stronger evidence for each question so that the chance of claiming statistical association when none exists is only 1% instead of $p < 5$
- This insures that, overall, no more than 5% false positive errors
  - Various methods guide how much stronger one should require evidence to be in different circumstances

Composite outcome

- Combine multiple outcomes into one index, and use that as a single outcome
- Multivariate → Univariate
  - Meld several outcomes into one question, by asking whether any function of any of them is associated with the exposure, e.g., treatment, and seeking a single yes or no

Outcome hierarchy

- Order outcomes from most to least important
- Crucial decisions in planning and analyzing the study are based on the primary outcome atop the hierarchy
Multiple outcome variables

- Requiring stronger evidence to claim an effect is often overly conservative
  - Also, it's hard to get folks to implement it sincerely
  - Somehow lots of 'butts' get into the interpretation
- Composite outcomes and outcome hierarchies require choices clinical researchers are not always comfortable making
  - Results may be hard to interpret
  - Composite outcome methods are being

Disclose “data snooping”

- Only pre-specified comparisons are to be

Effect of Systemic Matrix Metalloproteinase Inhibition on Periodontal Wound Repair: A Proof of Concept Trial

- **Background:** The adjunctive use of matrix metalloproteinase (MMP) inhibitors with scaling and root planing (SRP) promotes new attachment in patients with periodontal disease. This pilot study was designed to examine aspects of the biological response brought about by the MMP inhibitor low dose doxycycline (LDD) combined with access flap surgery (AFS) on the modulation of periodontal wound repair in patients with severe chronic periodontitis.
- **Methods:** Twenty-four subjects were enrolled into a 12-month, randomized, placebo-controlled, double-masked trial to evaluate clinical, biochemical, and microbial measures of disease in response to 6 months therapy of either placebo capsules + AFS or LDD (20 mg b.i.d.) + AFS. Clinical measures including probing depth (PD), clinical attachment levels (CAL), and bleeding on probing (BOP) as well as gingival crevicular fluid bone marker assessment (ICTP) and microbial DNA analysis (levels and proportions of 40 bacterial species) were performed at baseline and 3, 6, 9, and 12 months.
- **Results:** Patients treated with LDD + AFS showed more potent reductions in PD in surgically treated sites of >6 mm (P <0.05, 12 months). Furthermore, LDD + AFS resulted in greater reductions in ICTP levels compared to placebo + AFS. Rebounds in ICTP levels were noted when the drug was withdrawn. No statistical differences between the groups in mean counts were found for any pathogen tested.
- **Conclusions:** This pilot study suggests that LDD in combination with AFS may improve the response of surgical therapy in reducing probing depth in severe chronic periodontal disease. LDD administration also tends to reduce local periodontal bone resorption during drug administration. The use of LDD did not appear to contribute to any significant shifts in the microbiota beyond that of surgery alone. Gapski, Barr, Sarment, Layher, Socransky, Giannobile. *J Periodontol* 2004;75:441-452.

Concepts

- You need a testable, FINER research question
- Try to rule out bias, confounding, chance
- Consider multiple outcome measures and multiple predictors
- Disclose what you did [and why] with enough detail so others may replicate

Mark Twain: There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment of fact.
Thanks

... for your participation

Acknowledgement: I’m grateful to Peter B Imrey, PhD for sharing his previous version of this talk.