All polar bears are left handed
Common “misuse”, “nonuse” and “abuse” of statistical language and concepts in manuscripts

Part I

Lehana Thabane, PhD

Department of Anesthesia Research Meeting

September 22, 2010
Most scientists regarded the new streamlined peer-review process as ‘quite an improvement.’
Reading published research reports, textbooks, etc
Review of students' theses/manuscripts

“IT doesn’t matter that your thesis is on extra sensory perception, you can’t cite mindreading in your references.”
The issues...

- Misuse of statistical terms
- Problems in describing the methods
- Errors in reporting analysis results
- Errors in presenting data in tables and graphs
- Errors in interpreting results

Discuss how to avoid them or deal with them
Part I

 популярен

 ◊ Misuse of statistical terms

 ◊ Problems in describing the methods
Misuse of statistical terms
(Interquartile) range

• Replace “range is 10 to 20” by
  - “min to max is 10 to 20”

• Replace “interquartile range is 12,15” by
  “first and third quartiles are 12 and 15”;  
  - if the length is meant, compute it!

• A **range** is the **length** of the observed interval
• Cut points and groups:
  - The one Median \((M)\) cuts the data into 2 halves; 
    the lower half and the upper half

• The 2 Tertiles \((T_1, T_2)\) cut the data into 3 thirds:
  - lower third, middle third, upper third

• The 3 Quartiles \((Q_1, Q_2=M, Q_3)\) cut the data into 4 quarters:
  - lowest quarter, second quarter, third quarter, 
    highest quarter

• Same applies to quintiles, deciles, 
  percentiles, etc
Variance vs variation

• Make sure “variance” and “variation” are used properly
  ✓ **Variance** is the square of SD with squared units
  ✓ **Variation** is the general concept of variability

• Use “SD” to record variation rather than “SE”; even though SE < SD
  - SD has the same units as the data

• A square root of a variance is a standard deviation (SD)!
Study sample vs study population

• A “population” is ALL who meet the inclusion/exclusion criteria
  - The study group is those that get into the study; NOT the study population

• Your “sample” is not a population; i.e., avoid “patient population”
  - If it were a population, you would not need statistical inference!

• “Population Based” needs to show it is ALL in the region being studied.
Statistical significance vs clinical importance

- Save “Significance” for a statistical context, “Importance” for the content,
  - ie, clinically
  - Statistically significant and clinically important are two distinct concepts.

- Significance is a statement about an hypothesis; not part of it

- Replace
  - “insignificance” by “non significance”
  - “insignificant” by “not significant”
Multivariable vs multivariate...

“Multivariate” is the simultaneous analysis of 2 or more outcomes

✓ Clinically oriented or “traditional” outcomes
  ➤ Mortality
  ➤ Pain score
  ➤ Major morbidity
  - Cardiovascular, coagulation, cognitive,
  - gastrointestinal, immune, pulmonary, stress response

✓ “Nontraditional” outcomes
  ➤ Health-related quality-of-life
  ➤ Patient satisfaction
  ➤ Economic outcomes
“Multivariable” is any analysis that contains 2 or more independent variables in the analysis

✓ **Single outcome**: HRQoL

✓ **Independent variables**: gender, age, hypertension, weight, etc
• **Data** is a plural word;
  - use “are” rather than “is”, or “were” rather than “was”
  - Datum is singular

• Replace “an average” by “a mean”

• Avoid use of “average” choose “typical” so as not be confused with a statistical judgment
  - Do not use “, on average,” in your sentences
  - Do not use “an average patient”
“Control/experimental event rate” vs “Control/experimental event risk”

“The control event risk was 5%…” Or “the control event rate was 0.05…”

- A rate is not equivalent to a risk
- Both are ratios, but a risk is a proportion
- A rate has time units

Keep in mind…

- Risk (scale = [0,1])
- Proportion (scale = [0,1])
- Probability (scale = [0,1])
- Percentage (scale = [0,100])
- Rate (scale = [0, X] where X is a multiplier)

Other misuses...

• Replace “bilateral” by “two-tailed” test

• Replace “ranged” by “varied”, “range” by “vary”, or “ranging” by “varying” (verbs)

• Replace “range’ by “interval” (nouns)

• Replace “parameter” by “variable”
  - A parameter is a characteristic of a population distribution of a variable; not another name for a variable

• Insert “estimated” in front of “kappa” since you are reporting about a sample not the parameter in a population;
  - or use “kappa estimate”
Problems in describing the methods
Describing the Methods: misuse

- Replace “P-values < 0.05 were considered significant” with “the criterion for statistical significance was set at alpha = 0.05”
- “We used the Mann-Whitney test (or non-parametric tests) because of the small sample size…”
  - Note that non-parametric tests are generally less efficient than their parametric counterparts;
  - Their use should not be justified based on small sample sizes - because they would fair poorly in such cases;
  - The rationale for their use should be based on distributional assumptions.
Sample size: misuse

• Inappropriate justification for sample size
  - “So-and-so did a similar study with 6 patients and got statistical significance – ours uses 12 patients (double the size)”
  - “We did a similar pilot before and got it published”
  - “I don’t have the resources for a large multi-centre study”
  - “I have funding for only 10 patients!”
  - “I have limited SEED funding!”

• Provide details of sample size calculation
More problems with size…

• Not providing sufficient information to replicate the sample size (SS) calculation

• Not providing sufficient justification for parameters in the calculation

• Not having sample size procedure in alignment with primary research question or method of analysis

SS = f(8 ingredients)
✓ Study design
✓ Study hypothesis
✓ Nature of outcome
✓ Statistical methods
✓ Measure of variation
✓ Number of tails
✓ Important difference
✓ Significance / power
Reporting standards: Nonuse

• Not stating the guideline followed in reporting or not adhering to it (eg CONSORT)

• Not specifying how descriptive analyses of demographics, baseline characteristics or outcomes are reported
  - Mean (SD), median (min,max) or median (Q1,Q3) for continuous variables
  - Use number (percent) for categorical variables
• Failure to report
  - method for adjusting alpha for multiple comparisons (where applicable)
  - a reference for uncommon or complex methods

• Not stating how
  - Model assumptions (eg Normality, multicolinearity, etc) were assessed
  - Outliers or missing data were handled

• Not reporting the software (version) used for all analyses
In general, there is some confusion between ...

• **Secondary analyses**
  - relates to secondary outcomes

• **Subgroup analyses**
  - Relates to comparing different groups of patients on primary/secondary outcomes

• **Sensitivity analyses**
Sensitivity analyses: assessing the robustness of the main results to...

• Different methods of handling missing data

• Different methods of analysis (different assumptions)

• Outliers (analysis with and without outliers)

• Different definitions of outcomes (ie different cut-off points for binary outcomes)

• Adjusting for possible baseline imbalance in RCTs
  - A.k.a. “adjusted analysis”
  - Often incorrectly referred to as “secondary analysis”
Thank You!

thabanL@mcmaster.ca