Advanced Analysis Methods
Issues to Consider in the Analysis

Workshop Day 2 AM
November 20, 2007

George A Wells
University of Ottawa

Issues to Consider in the Analysis

- **Unit of analysis**
- **Intention to treat**
- **Multiple arm trials**
- **Continuous data: variations in reporting**
- **Continuous data: changes from baseline**

Issues to Consider in the Analysis
Unit of Analysis

- usually the number of observations analyzed matches the number of units randomized
- must consider if each participant is:
  - only one member of a group randomized (eg. cluster)
  - undergoes more than one treatment (eg. cross-over)
  - yields multiple measurements (eg. repeated measurements, recurring events, measurements on several body parts)

Issues to Consider in the Analysis
Unit of Analysis

Repeated Measurements

- define several outcomes based on different follow-up times and perform separate analysis (eg. short, medium, long term follow-up)
- select single time point and analyze only this for trials in which it is presented
- select longest follow-up from each trial (may induce lack of consistency leading to heterogeneity)
- seek individual patient data and create a single outcome for each participant (eg. rate of change)

Issues to Consider in the Analysis
Unit of Analysis

Recurring Events

- preferable to consider the number of times these events occur … 'count data'
- rare vs common events
  - counts of rare events are known as 'Poisson data' and analyses focus on rates (relating the counts to amount of time during which they could have occurred)
  - counts of more common events may often be treated as continuous data

Issues to Consider in the Analysis
Unit of Analysis

Multiple Treatment Attempts

- if use participant as 'denominator' and not treatment attempts then no issue
- otherwise it is a cluster trial with participant as the cluster

Issues to Consider in the Analysis
Unit of Analysis

Multiple Body Parts

- multiple body parts receive same treatment and number of body parts used as denominator in analysis
  - cluster trial with participant as the cluster
- multiple parts are randomized to different treatments
  - cross-over trial

Issues to Consider in the Analysis
Unit of Analysis

Multiple Intervention Groups

- if a trial has more than 2 intervention groups then problems arise if the same group is included twice in same meta-analysis
  - Example: re-using the control group

Issues to Consider in the Analysis
Unit of Analysis

Outcomes applicable to a subset of participants

- **Example: fertility studies with miscarriage as an outcome**
  - **proportion of randomized patients that miscarry**
    versus
    proportion of clinical pregnancies that miscarry

Issues to Consider in the Analysis

- **Unit of analysis**
- **Intention to treat**
- **Multiple arm trials**
- **Continuous data: variations in reporting**
- **Continuous data: changes from baseline**

Issues to Consider in the Analysis
Intention to Treat

- Two issues involved
  - participants should be analyzed in the groups to which they were randomized regardless of treatment received
  - outcomes should be included for all participants regardless of whether outcomes were collected
- ITT preferred (some exceptions; eg non-inferiority trial)
- if dropout high or different across treatment groups then analysis plan may dictate study excluded from meta-analysis
- should attempt to extract data from papers as close to ITT principles as possible

Issues to Consider in the Analysis
Intention to Treat

- Do trials present ITT analyses
  - some trials do not report ITT analyses but only per protocol analyses
  - many trials report ITT analysis taking into consideration only the issue of 'analyzed to treatment randomized'
    - when contact lost with participants difficult to include all participants 'regardless of whether outcomes were collected'
  - in some cases can create ITT analysis from information in paper or from additional information from author

Issues to Consider in the Analysis
Intention to Treat

Dichotomous data

- basic data needed
  - number assessed and number incurring event in each group

Issues to Consider in the Analysis
Intention to Treat

Dichotomous data (RR, OR, RD)

	Event	No event		
Experimental	a	b	n_e	$P_e = a/n_e$
Control	c	d	n_c	$P_c = c/n_c$

$$RR = P_e / P_c = \frac{a(c+d)}{(a+b)c}$$

$$OR = \frac{P_e}{1-P_e} \Big/ \frac{P_c}{1-P_c} = \frac{ad}{bc}$$

$$RD = P_e - P_c = \frac{a}{a+b} - \frac{c}{c+d}$$

Basic Data: a / n_e c / n_c

Issues to Consider in the Analysis
Intention to Treat

Dichotomous data

- no consensus on best way to handle in the analysis 'participants providing no outcome data'
  - include data only on those with known results (impact on results should be considered and will depend on degree of missing, event frequency and effect size; consider as source of heterogeneity)
  - impute outcomes for missing participants: inflate numerators and denominators by same percentage so risks are unchanged but sample size reflects initial numbers randomized
  - assume missing participants all experienced the event or all did not experience the event ; do sensitivity analysis by varying the study groups on these two assumptions

Issues to Consider in the Analysis
Intention to Treat

Continuous data

- basic data needed
  - number assessed, mean and standard deviation (sd) in each group

Issues to Consider in the Analysis
Intention to Treat

Continuous Data (MD, SMD)

	number	mean	standard deviation
Experimental	n_e	$\bar{X}_e$	s_e
Control	n_c	$\bar{X}_c$	s_c

Mean difference (MD) : $\bar{x}_e - \bar{x}_c$

SMD : $d = f\,\dfrac{\bar{x}_e - \bar{x}_c}{s}$

where :

$$s = \sqrt{\frac{(n_e - 1)s_e^2 + (n_c - 1)s_c^2}{n_e + n_c - 2}}$$

$$f = \frac{4(n_e + n_c - 2) - 4}{4(n_e + n_c - 2) - 1}$$

Basic Data n_e $\bar{X}_e$ s_e n_c $\bar{X}_c$ s_c

Issues to Consider in the Analysis
Intention to Treat

Continuous data

- possible ways to handle these participants in analysis
  - include data only on those with known results (impact on results should be considered and will depend on degree of missing, inherent variability and effect size; consider as source of heterogeneity)
  - impute outcomes for missing participants:
    - simply inflate sample size to the number randomized (big assumption that those missing have same mean and sd as those in the group they were allocated to);
    - in some cases may be able to use standard approaches (eg. last observation carried forward using baseline values etc.)

Issues to Consider in the Analysis

- **Unit of analysis**
- **Intention to treat**
- **Multiple arm trials**
- **Continuous data: variations in reporting**
- **Continuous data: changes from baseline**

Issues to Consider in the Analysis
Multiple Arm Treatments

- **example: Treatment groups A, B, Control**
  - **want to include in same analysis the pairwise comparisons**
    A vs Control
    B vs Control
- **if you want to reuse the Control in this way then you must 'adjust' so that the 'impact' of the Control group is not 'twice' it should be**
- **three approaches**

Issues to Consider in the Analysis
Multiple Arm Treatments

Approach 1

- include only one pairwise comparison choosing the most 'standard' experimental group (A or B)

- disadvantage
  - information from other experimental group not used

Issues to Consider in the Analysis
Multiple Arm Treatments

Approach 2

- Include each pairwise comparison separately but with the repeated control group divided out evenly among the comparisons

- Dichotomous data:
  - basic data in Control group: a/n
  - assume in Control 1: n/2 and a/2
    Control 2: n/2 and a/2

- Continuous data:
  - basic data in Control group: n, mean, sd
  - assume in Control 1: n/2
    Control 2: n/2

- Disadvantages:
  - decreased sample size in each control group
  - compromises some assumptions underlying a random effects model

Issues to Consider in the Analysis
Multiple Arm Treatments

Approach 3

- Combined all experimental groups into one group and all control groups into one group
- Dichotomous data:
  - basic data: number assessed and number incurring event
  - simply sum these corresponding numbers over the experimental groups; sum the corresponding numbers for the control groups
- Continuous data:
  - basic data: number assessed (n), mean and standard deviation (sd) in each group
  - weighted mean, pooled standard deviation and sum of number assessed over the experimental groups; similarly for the control groups
- Disadvantages:
  - lose individual pairwise comparisons
  - possibly combining 'unlike' groups

Issues to Consider in the Analysis

- Unit of analysis
- Intention to treat
- Multiple arm trials
- Continuous data: variations in reporting
- Continuous data: changes from baseline

Issues to Consider in the Analysis
Continuous Data: variations in reporting

- **Average**
  - mean (m)
  - median (M)
- **Variation**
  - standard deviation (SD)
  - standard error (SE)
  - confidence interval (CI)
  - interquartile range (IQR)
  - range (R)
- **Basic data:**
  - number assessed, mean and standard deviation (SD) in each group

Issues to Consider in the Analysis
Continuous Data: variations in reporting

Standard Error (SE)

$$SD = (SE)\sqrt{n}$$

Issues to Consider in the Analysis
Continuous Data: variations in reporting

Confidence Interval (CI)

- CI = (lcl, ucl); check for symmetry
- for large n (n>100) use
  3.92 (95% CI) 3.29 (90%CI) 5.15 (99%CI)
- for small n use values t-distribution with n-1 df

$$SD = (ucl - lcl) / 3.92x\sqrt{n}$$

Issues to Consider in the Analysis
Continuous Data: variations in reporting

t-test for MD (t)

$$t = MD / \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} SD$$

$$SD = MD / t \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Issues to Consider in the Analysis
Continuous Data: variations in reporting

Confidence Interval for MD

- CI = (lcl, ucl); check for symmetry
- for large n_1, n_2 use
  3.92 (95% CI) 3.29 (90%CI) 5.15 (99%CI)
- for small n_1, n_2 use values t-distribution with $n_1 + n_2 - 2$ df

$$SD = (ucl - lcl) / 3.92x \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Issues to Consider in the Analysis
Continuous Data: variations in reporting

Exact p-value for t-test of MD

- use 't-table' corresponding to p-value and $n_1 + n_2 - 2$ df
- transform t as described earlier

$$t = MD / \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} SD$$

$$SD = MD / t \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Issues to Consider in the Analysis
Continuous Data: variations in reporting

Interquartile range (IQR)

- if n is reasonably large and distribution of the outcome is similar to a normal distribution, then IQR = 1.35 SD

$$SD = IQR / 1.35$$

Issues to Consider in the Analysis

- Unit of analysis
- Intention to treat
- Multiple arm trials
- Continuous data: variations in reporting
- Continuous data: changes from baseline

Issues to Consider in the Analysis
Continuous Data: changes from baseline

Relating two numbers (x,y)

- incremental change — x-y
- proportional change — (x-y)/y
- percent change — 100 (x-y)/y
- proportion (of baseline) — x/y
- percent (of baseline) — 100 x/y

Issues to Consider in the Analysis
Continuous Data: changes from baseline

	Experimental	Control
Baseline	n_{e1} , x_{e1} , s_{e1}	n_{c1} , x_{c1} , s_{c1}
End of Study	n_{e2} , x_{e2} , s_{e2}	n_{c2} , x_{c2} , s_{c2}
Change from baseline	n_e , x_e , s_e	n_c , x_c , s_c

$x_e = x_{e2} - x_{e1}$

$$S_e = \sqrt{S_{e1}^2 + S_{e2}^2 - 2\rho S_{e1} S_{e2}}$$

$x_c = x_{c2} - x_{c1}$

$$S_c = \sqrt{S_{c1}^2 + S_{c2}^2 - 2\rho S_{c1} S_{c2}}$$

Issues to Consider in the Analysis

- **Unit of analysis**
- **Intention to treat**
- **Multiple arm trials**
- **Continuous data: variations in reporting**
- **Continuous data: changes from baseline**

Advanced Analysis Methods
Issues to Consider in the Analysis

Workshop Day 2 AM
November 20, 2007