Meta-analysis of count data

Peter Herbison
Department of Preventive and Social Medicine
University of Otago

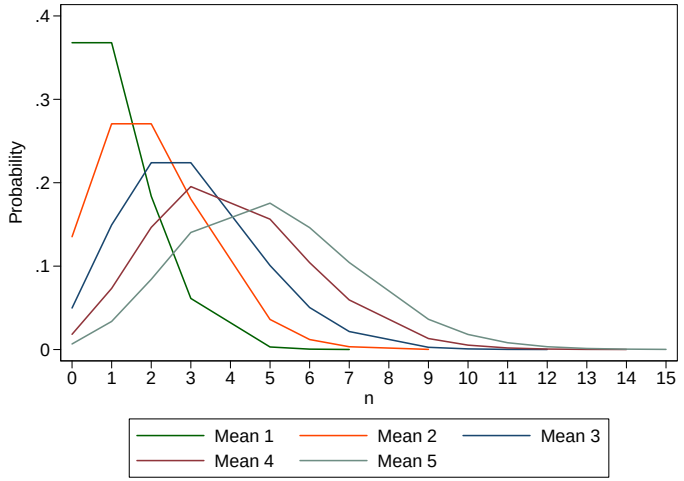
What is count data

- Data that can only take the non-negative integer values 0, 1, 2, 3, …….
- Examples in RCTs
  - Episodes of exacerbation of asthma
  - Falls
  - Number of incontinence episodes
- Time scale can vary from the whole study period to the last few (pick your time period).

Complicated by

- Different exposure times
  - Withdrawals
  - Deaths
  - Different diary periods
    - Sometimes people only fill out some days of the diary as well as the periods being different
- Counts with a large mean can be treated as normally distributed

What is large?



Analysis as rates

- Count number of events in each arm
- Calculate person years at risk in each arm
- Calculate rates in each arm
- Divide to get a rate ratio
- Can also be done by poisson regression family
  - Poisson, negative binomial, zero inflated poisson
  - Assumes constant underlying risk

Meta-analysis of rate ratios

- Formula for calculation in Handbook
- Straightforward using generic inverse variance
- In addition to the one reference in the Handbook:
  - Guevara et al. Meta-analytic methods for pooling rates when follow up duration varies: a case study. BMC Medical Research Methodology 2004;4:17

Dichotimizing

- Change data so that it is those who have one or more events versus those who have none
- Get a 2x2 table
- Good if the purpose of the treatment is to prevent events rather than reduce the number of them
- Meta-analyse as usual

Treat as continuous

- May not be too bad
- Handbook says beware of skewness

Time to event

- Usually time to first event but repeated events can be used
- Analyse by survival analysis (e.g. Cox's proportional hazards regression)
- Ends with hazard ratios
- Meta-analyse using generic inverse variance

Not so helpful

- Many people look at the distribution and think that the only distribution that exists is the normal distribution and counts are not normally distributed so they analyse the data with non-parametric statistics
- Not so useful for meta-analyses

Problems with primary studies

- Can be difficult to define events
  - Exacerbations of asthma/COPD
  - Osteoporosis fractures
- Need to take follow-up time into account in analysis
- Failure to account for overdispersion

So what is the problem?

- Too many choices
- All reasonable things to do
- So what if studies choose different methods of analysis?

Rates and dichotomizing

- From handbook
  - "In a randomised trial, rate ratios may often be very similar to relative risks obtained after dichotomizing the participants, since the average period of follow-up should be similar in all intervention groups. Rate ratios and relative risks will differ, however, if an intervention affects the likelihood of some participants experiencing multiple events".

Combining different analyses

- So is it reasonable to combine different methods of analysing the data?
- A colleague who works on the falls review and I have been thinking about this for a while
- Falls studies are quite fragmented partly because of the different methods of analysis

Simulation

- Simulated data sets with different characteristics
- Analysed them many ways
  - Rate ratio
  - Dichotomized
  - Poisson regression (allowing for duration)
  - Negative binomial (allowing for duration)
  - Time to first event
  - Means
  - Medians

Data sets

- Two groups (size normal(100,2))
- Low mean group (Poisson 0.2, 0.15)
- Medium mean group (2, 1.5)
- High mean group (7, 5)
- Overdispersion built in by 0%, 20% and 40% from a distribution with a higher mean
- 20% not in the study for the full time (uniform over the follow up)

Low mean no overdispersion

- Rate ratio = 0.7977

	Mean difference	SD
binary	-0.0121	0.0839
poisson	0.0000	0.0001
neg binom	0.0006	0.0089
hazard	-0.0058	0.1072
mean	0.0028	0.0320
median	Not possible	

Low mean some overdispersion

- Rate ratio = 0.7521

	Mean difference	SD
binary	-0.0376	0.0955
poisson	0.0000	0.0001
neg binom	0.0019	0.0124
hazard	-0.0252	0.1114
mean	0.0008	0.0304
median	Not possible	

Low mean high overdispersion

- Rate ratio = 0.7205

	Mean difference	SD
binary	-0.0417	0.0852
poisson	0.0000	0.0001
neg binom	0.0015	0.0122
hazard	-0.0170	0.0959
mean	-0.0008	0.0293
median	Not possible	

Medium mean no overdispersion

- Rate ratio = 0.7578

	Mean difference	SD
binary	-0.1433	0.0723
poisson	0.0000	0.0001
neg binom	0.0024	0.0069
hazard	-0.0616	0.1081
mean	0.0008	0.0283
median	Not possible	

Medium mean some overdispersion

- Rate ratio = 0.7789

	Mean difference	SD
binary	-0.1362	0.0752
poisson	0.0000	0.0001
neg binom	0.0032	0.0108
hazard	-0.0499	0.1211
mean	0.0008	0.0291
median	-0.0758	0.1770

Medium mean high overdispersion

- Rate ratio = 0.7979

	Mean difference	SD
binary	-0.1318	0.0739
poisson	0.0000	0.0001
neg binom	0.0030	0.0126
hazard	-0.0508	0.1236
mean	0.0008	0.0299
median	-0.0743	0.1400

High mean no overdispersion

- Rate ratio = 0.7154

	Mean difference	SD
binary	-0.2783	0.0493
poisson	0.0000	0.0001
neg binom	0.0096	0.0211
hazard	-0.2769	0.1400
mean	0.0000	0.0272
median	0.0064	0.0562

High mean some overdispersion

- Rate ratio = 0.7101

	Mean difference	SD
binary	-0.2846	0.0490
poisson	0.0000	0.0001
neg binom	0.0125	0.0292
hazard	-0.2841	0.1522
mean	-0.0001	0.0271
median	-0.0005	0.0588

High mean high overdispersion

- Rate ratio = 0.7068

	Mean difference	SD
binary	-0.2891	0.0504
poisson	0.0000	0.0001
neg binom	0.0148	0.0309
hazard	-0.2912	0.1552
mean	0.0000	0.0271
median	0.0018	0.0503

Conclusions 1

- When you have a low mean you should be able to combine almost anything
- As the mean increases then dichotomizing events increasingly underestimates treatment effects
- Time to first event underestimates but to a lesser extent than dichotomizing
  - Allowing for multiple events may help

Conclusions 2

- Adjusting for overdispersion by using negative binomial has only a small effect even for a quite a bit of overdispersion
  - In spite of neg bin being a better fit
- Means (at least ratio of means) is surprisingly good
- Ratio of medians is astonishing, but inefficient
  - Problems with SE