Meta-analysis of ordinal data

Steff Lewis

Edinburgh MRC clinical trials methodology hub

(with thanks to Izzy Butcher, Gillian McHugh and Jim Weir for examples)

Examples of ordinal scales in stroke

- **Modified Rankin Scale**

Score	Description
0	No symptoms at all
1	No significant disability despite symptoms; able to carry out all usual duties and activities
2	Slight disability; unable to carry out all previous activities, but able to look after own affairs without assistance
3	Moderate disability; requiring some help, but able to walk without assistance
4	Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance
5	Severe disability; bedridden, incontinent and requiring constant nursing care and attention
6	Dead

- Barthel Index

Feeding	0 = unable; 1 = needs help; 2 = independent
Bathing	0 = dependent; 1 = independent
Grooming	0 = needs help; 1 = independent
Dressing	0 = dependent; 1 = needs help; 2 = independent
Bowels	0 = incontinent; 1 = occasional accident; 2 = continent
Bladder	0 = incontinent; 1 = occasional accident; 2 = continent
Toilet Use	0 = dependent; 1 = needs some help; 2 = independent
Transfers (bed to chair, and back)	0 = unable; 1 = major help; 2 = minor help; 3 = independent
Mobility (on level surfaces)	0 = immobile or < 50 yards; 1 = wheelchair independent, > 50 yards; 2 = walks with help > 50 yards; 3 = independent > 50 yards
Stairs	0 = unable; 1 = needs help; 2 = independent
	TOTAL (0–20)

How common are ordinal data?

- Cochrane stroke group has 118 full reviews of the effectiveness of interventions (12 Jan 2010).
- Approx 2/3 have an ordinal outcome measure.
- None are analysed as ordinal data.
- They either dichotomise [approx 3/4] or treat as continuous [approx 1/4].

What the Handbook says……

9.4.7 Meta-analysis of ordinal outcomes and measurement scales

Ordinal and measurement scale outcomes are most commonly meta-analysed as dichotomous data or continuous data depending on the way that the study authors performed the original analyses

How common are ordinal outcomes in other review groups?

- Does anyone know of any ordinal analyses in Cochrane that use methods other than those available in Revman?

What's wrong with analysing ordinal data as if they are binary?

Individuals who fall close to, but on different sides of the cut-point, will be assumed by the analysis to be different, yet they are likely to be similar.

- **Modified Rankin Scale**

Score	Description
0	No symptoms at all
1	No significant disability despite symptoms; able to carry out all usual duties and activities
2	Slight disability; unable to carry out all previous activities, but able to look after own affairs without assistance
3	Moderate disability; requiring some help, but able to walk without assistance
4	Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance
5	Severe disability; bedridden, incontinent and requiring constant nursing care and attention
6	Dead

Individuals who improve, but don't improve past the cut-point won't be counted as improvers in the analysis.

- **Modified Rankin Scale**

Score	Description
0	No symptoms at all
1	No significant disability despite symptoms; able to carry out all usual duties and activities
2	Slight disability; unable to carry out all previous activities, but able to look after own affairs without assistance
3	Moderate disability; requiring some help, but able to walk without assistance
4	Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance
5	Severe disability; bedridden, incontinent and requiring constant nursing care and attention
6	Dead

It is throwing away information

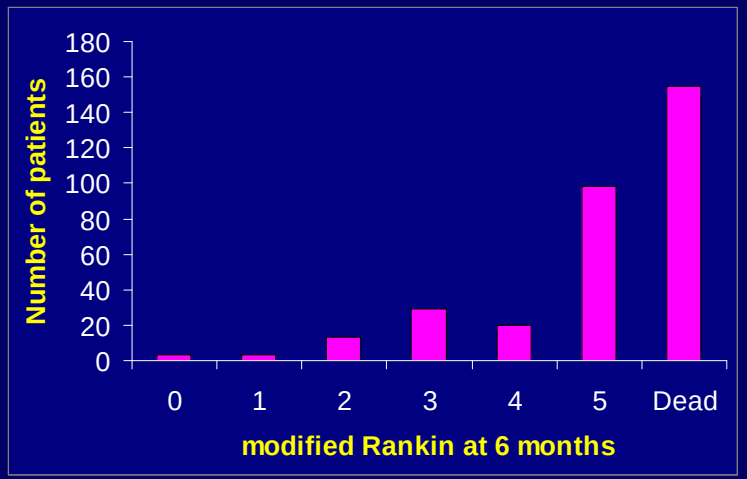
- In individual studies, for continuous data:
  - The loss of power in dichotomising continuous data at the mean is equivalent to throwing away a third of the data.
  - Dichotomising away from the mean is even worse.
  - Cohen J. *Appl Psychol Meas* 1983;**7**:249.
- The same concepts are true of ordinal data.
  - Re-analysis of ordinal data in individual stroke trials has shown that sample sizes could be around 30% smaller if data were analysed using the full ordinal scale rather than by dichotomising [OAST 2008].
  - Similar results occur in head injury (IMPACT team)

What's wrong with analysing ordinal data as if they are continuous?

(using standard methods in Revman)

- There may be nonparametric methods that use rankings that are OK (although may not give good summary estimates for meta-analysis)

The data may not be Normally distributed



FOOD trial – PEG vs NG feeding tubes in stroke patients

May not be a linear scale so change from 1 to 2 is not the same as 2 to 3.

- **Modified Rankin Scale**

Score	Description
0	No symptoms at all
1	No significant disability despite symptoms; able to carry out all usual duties and activities
2	Slight disability; unable to carry out all previous activities, but able to look after own affairs without assistance
3	Moderate disability; requiring some help, but able to walk without assistance
4	Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance
5	Severe disability; bedridden, incontinent and requiring constant nursing care and attention
6	Dead

So what can we do instead?

- Proportional odds modelling
  - Makes no distributional assumptions about the outcome

Proportional odds model

- Proportional odds model assumes there is an equal odds ratio for all dichotomies of the data.
- The odds ratio calculated from the proportional odds model can be interpreted as the odds of success on the experimental intervention relative to control, irrespective of how the ordered categories might be divided into success or failure.

SAPHIR trial

Glasgow outcome scale, for those with and without subarachnoid haemorrhage

	Dead/Veg	Severe	Moderate	Good
No	88	73	100	247
Yes	147	65	73	113

Dichotomies

	Good	Moderate	Severe	Veg/Dead
No	88	73	100	247
Yes	147	65	73	113

Odds ratio 2.80

Dichotomies

	Good	Moderate	Severe	Veg/Dead
No	88	73	100	247
Yes	147	65	73	113

Odds ratio 2.46

Dichotomies

	Good	Moderate	Severe	Veg/Dead
No	88	73	100	247
Yes	147	65	73	113

Odds ratio 2.39

Odds Ratios

	SAPHIR	TINT
Not Good	2.80	2.28
Unfavourable	2.46	3.10
Dead/Veg	2.39	3.48
Proportional odds model	2.51	2.73

Pitfalls, etc

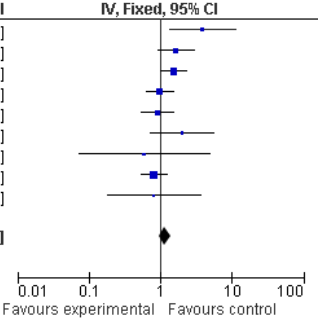
- IMPACT head injury investigators have found that the proportional odds assumption mostly holds in their trial data.
- They say even if the data deviate considerably from proportional odds, it still gives a useful summary.
- However, it will hide 'kill or cure' effects if used without any other summary measures.

Thrombolysis (tPA) for acute ischaemic stroke

– Death during follow up

Study or Subgroup	Thrombolysis Events	Thrombolysis Total	Control Events	Control Total	Weight	Odds Ratio IV, Fixed, 95% CI
ATLANTIS A 2000	16	71	5	71	3.3%	3.84 [1.32, 11.15]
ATLANTIS B 1999	33	307	21	306	11.5%	1.63 [0.92, 2.90]
ECASS 1995	69	313	48	307	22.5%	1.53 [1.01, 2.29]
ECASS 2 1998	43	409	42	391	18.5%	0.98 [0.62, 1.53]
ECASS 3 2008	32	418	34	403	14.8%	0.90 [0.54, 1.49]
EPITHET 2008	13	52	7	49	3.6%	2.00 [0.72, 5.53]
Mori 1992	2	19	2	12	0.8%	0.59 [0.07, 4.85]
NINDS 1995	54	312	64	312	23.2%	0.81 [0.54, 1.21]
Wang 2003	5	67	3	33	1.7%	0.81 [0.18, 3.60]
Total (95% CI)		**1968**		**1884**	**100.0%**	**1.16 [0.95, 1.40]**
Total events	267		226			

Heterogeneity: Chi² = 14.28, df = 8 (P = 0.07); I² = 44%
Test for overall effect: Z = 1.46 (P = 0.14)



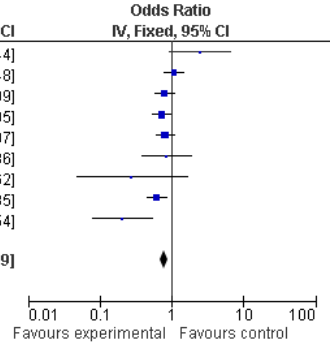
From Wardlaw JM et al. *Cochrane Database of Systematic Reviews* 2009, Issue 4. Art. No.: CD000213.(Only studies that report both death, and death and dependency included)

Thrombolysis (tPA) for acute ischaemic stroke

– Death or dependency during follow up

Study or Subgroup	Thrombolysis Events	Thrombolysis Total	Control Events	Control Total	Weight	Odds Ratio IV, Fixed, 95% CI
ATLANTIS A 2000	64	71	56	71	1.8%	2.45 [0.93, 6.44]
ATLANTIS B 1999	141	307	135	306	16.8%	1.08 [0.78, 1.48]
ECASS 1995	171	313	185	307	16.7%	0.79 [0.58, 1.09]
ECASS 2 1998	187	409	211	391	22.0%	0.72 [0.54, 0.95]
ECASS 3 2008	140	418	155	403	20.9%	0.81 [0.61, 1.07]
EPITHET 2008	28	51	29	49	2.7%	0.84 [0.38, 1.86]
Mori 1992	11	19	10	12	0.5%	0.28 [0.05, 1.62]
NINDS 1995	155	312	192	312	16.8%	0.62 [0.45, 0.85]
Wang 2003	29	67	26	33	1.8%	0.21 [0.08, 0.54]
Total (95% CI)		**1967**		**1884**	**100.0%**	**0.78 [0.68, 0.89]**
Total events	926		999			

Heterogeneity: Chi² = 20.50, df = 8 (P = 0.009); I² = 61%
Test for overall effect: Z = 3.75 (P = 0.0002)



From Wardlaw JM et al. *Cochrane Database of Systematic Reviews* 2009, Issue 4. Art. No.: CD000213.(Only studies that report both death, and death and dependency included)

An example of a proportional odds meta-analysis

Data of the form…

Glasgow outcome scale, for those with and without active treatment

	Dead/Veg 1	Severe 2	Moderate 3	Good 4
Trt = 0	*n*	*n*	*n*	*n*
Trt = 1	*n*	*n*	*n*	*n*

SAS code

```
proc sort;
   by trial;
proc logistic order=internal;
   class treatment (param=ref ref='0');
   model ordscale(descending) = treatment;
   weight n;
   by trial;
run;
```

SAS output

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 4	1	-0.4089	0.0841	23.6336	<.0001
Intercept 3	1	0.7979	0.0881	81.9773	<.0001
Intercept 2	1	2.0064	0.1147	305.9194	<.0001
treatment 1	1	-1.1476	0.1544	55.2477	<.0001

- In Revman, use Generic inverse variance

for head injury

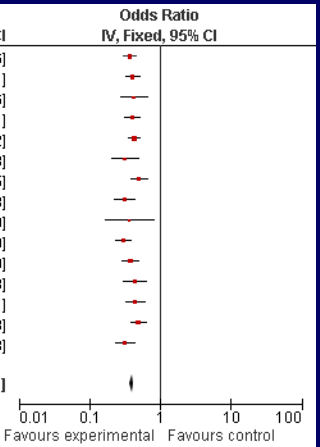
Text of Review | 1.1 Glasgow outcome...

Comparison: 1 Rock Music versus no treatment, Outcome: 1.1 Glasgow outcome score

Study or Subgroup	log[Odds Ratio]	SE	Weight	Odds Ratio IV, Fixed, 95% CI
AC/DC	-0.9956	0.1137	10.4%	0.37 [0.30, 0.46]
Aerosmith	-0.906	0.1172	9.8%	0.40 [0.32, 0.51]
Bon Jovi	-0.8582	0.2224	2.7%	0.42 [0.27, 0.66]
Guns 'N Roses	-0.9152	0.1233	8.8%	0.40 [0.31, 0.51]
Led Zeppelin	-0.8346	0.0961	14.6%	0.43 [0.36, 0.52]
Metallica	-1.1469	0.2143	2.9%	0.32 [0.21, 0.48]
Nirvana	-0.6966	0.1343	7.5%	0.50 [0.38, 0.65]
Pink Floyd	-1.1651	0.1623	5.1%	0.31 [0.23, 0.43]
Queen	-1.0054	0.3991	0.8%	0.37 [0.17, 0.80]
Rush	-1.2078	0.1343	7.5%	0.30 [0.23, 0.39]
Soundgarden	-0.9687	0.1377	7.1%	0.38 [0.29, 0.50]
Stone Temple Pilots	-0.8283	0.1874	3.8%	0.44 [0.30, 0.63]
U2	-0.8238	0.1641	5.0%	0.44 [0.32, 0.61]
Van Halen	-0.7187	0.1269	8.3%	0.49 [0.38, 0.63]
ZZ Top	-1.1476	0.1544	5.6%	0.32 [0.23, 0.43]
Total (95% CI)			100.0%	0.40 [0.37, 0.42]
Heterogeneity: Chi² = 17.48, df = 14 (P = 0.23); I² = 20%				
Test for overall effect: Z = 25.30 (P < 0.00001)				

Study or Subgroup	log[Odds Ratio]	SE	Weight	Odds Ratio IV, Fixed, 95% CI
AC/DC	-0.9956	0.1137	10.4%	0.37 [0.30, 0.46]
Aerosmith	-0.906	0.1172	9.8%	0.40 [0.32, 0.51]
Bon Jovi	-0.8582	0.2224	2.7%	0.42 [0.27, 0.66]
Guns 'N Roses	-0.9152	0.1233	8.8%	0.40 [0.31, 0.51]
Led Zeppelin	-0.8346	0.0961	14.6%	0.43 [0.36, 0.52]
Metallica	-1.1469	0.2143	2.9%	0.32 [0.21, 0.48]
Nirvana	-0.6966	0.1343	7.5%	0.50 [0.38, 0.65]
Pink Floyd	-1.1651	0.1623	5.1%	0.31 [0.23, 0.43]
Queen	-1.0054	0.3991	0.8%	0.37 [0.17, 0.80]
Rush	-1.2078	0.1343	7.5%	0.30 [0.23, 0.39]
Soundgarden	-0.9687	0.1377	7.1%	0.38 [0.29, 0.50]
Stone Temple Pilots	-0.8283	0.1874	3.8%	0.44 [0.30, 0.63]
U2	-0.8238	0.1641	5.0%	0.44 [0.32, 0.61]
Van Halen	-0.7187	0.1269	8.3%	0.49 [0.38, 0.63]
ZZ Top	-1.1476	0.1544	5.6%	0.32 [0.23, 0.43]
Total (95% CI)			**100.0%**	**0.40 [0.37, 0.42]**

Heterogeneity: Chi² = 17.48, df = 14 (P = 0.23); I² = 20%
Test for overall effect: Z = 25.30 (P < 0.00001)



Collecting data

- You need the numbers of patients in each category of the ordinal scale for each intervention group if the proportional odds ratio method will be used.
- Full data probably more likely for shorter scales and more recent papers??

Gøtzsche paper

Table 1 Proportion of trials with optimal and usable reporting for meta-analysis (see definitions in text). Test for positive trend

	1966–77	*1978–82*	*1983–87*	*1988–97*	*Trend (p value)*
Global evaluation					
No of trials	35	31	33	28	
Optimal reporting (%)	43	52	33	36	0.32
Usable reporting (%)	71	81	67	57	0.15
Pain					
No of trials	16	27	25	30	
Optimal reporting (%)	13	15	28	47	0.003
Usable reporting (%)	50	30	48	63	0.09

Optimal reporting: original ordered categories (but various scales included). For pain on VAS, mean and SD were accepted.

Gøtzsche paper

Table 2 Method of data presentation in 144 reports of trials comparing two non-steroidal

Method of data presentation	*Global evaluation*	*Pain*
Usable reporting		
Numbers of patients in original ordered categories	52	7
Numbers in reduced ordered categories, at least 3	12	5
Numbers in only 2 (reduced) ordered categories	14	5
Mean and SD, before and after treatment	4	26
Mean and SD before, mean after treatment	1	1
Mean and SD after treatment	4	1
Mean change and SD for change	1	2
Unusable reporting		
Mean before and after treatment, no SD	8	26
Mean after treatment, no SD	6	
Mean change, no SD	5	6
Percentage change, no SD		1
Median and IQ* range, before and after treatment		
Median change and IQ range		
Geometric mean percentage change, no SD		3
Median change, no percentiles or SD		2
Median before and after treatment, no centiles		
No data shown	20	13
Total number of trials reporting the variable	127	98

*IQ = Interquartile.

ECASS1 (1995)

Table 3.—Results*

End Points	Intention-to-Treat Population		Target Population	
	Placebo	rt-PA	Placebo	rt-PA
Primary				
Barthel Index†				
Median score	75	85	80	90
P	.99		.16	
Modified Rankin Scale†				
Median score	3	3	3	2
P	.41		.035	
Secondary				
Scandinavian Stroke Scale, long-term score at day 90±14†				
Median score (maximum score=58)	36	39	37	43
P	.54		.04	
Combined Barthel Index/ Rankin Scale at day 90±14†‡				
Median score	90	97.5	90	100
P	.003		<.001	
Mortality at day 30, %§	12.7	17.9	11.7	14.6
P	.08		.36	
Relative risk (95% confidence interval)	1.22 (1.02-1.45)		1.17 (0.95-1.46)	

*rt-PA indicates recombinant tissue plasminogen activator.
†Wilcoxon test.
‡Survivors only.
§Fisher's exact test.

ECASS 1 text:

- "In the ITT analysis 29.3% of patients in the placebo arm and 35.7% of the rt-PA treated patients had RS scores better than 2 at 90 days (Table 3)"

NINDS (1995)

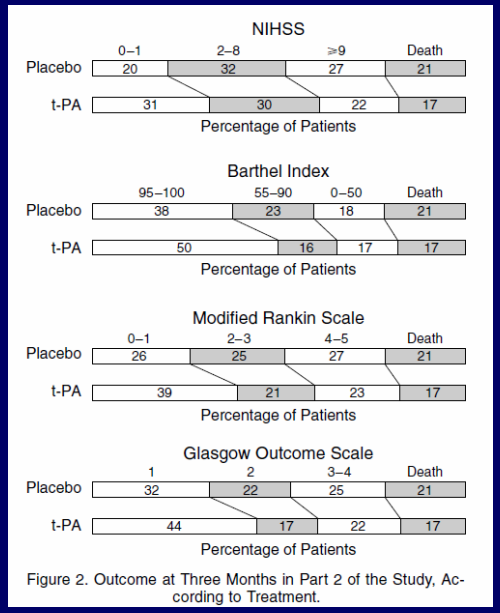


Figure 2. Outcome at Three Months in Part 2 of the Study, According to Treatment.

ECASS3 (2008)

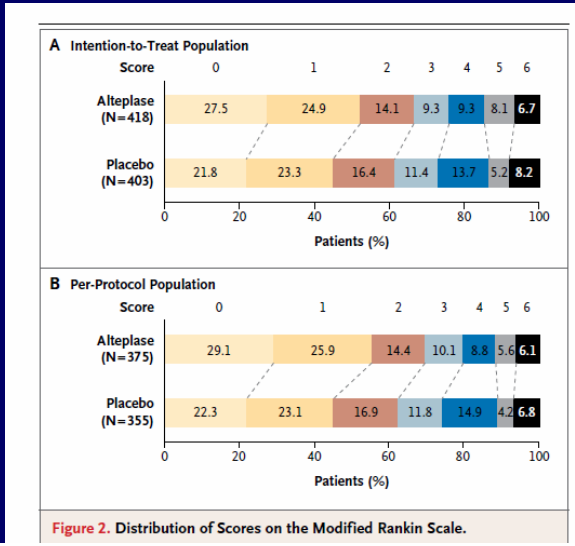


Figure 2. Distribution of Scores on the Modified Rankin Scale.

You could mix binary and ordinal data…

- Reminder: The odds ratio calculated from the proportional odds model can be interpreted as the odds of success on the experimental intervention relative to control, irrespective of how the ordered categories might be divided into success or failure.
- If proportional odds holds, you could combine:
  - The original Rankin scale in 7 categories
  - A summarised Rankin scale in 4 categories
  - Binary data where the scale has been split at 0-2 vs 3-6
  - Dead vs alive (category 6 on the scale vs 0-5).

Mixing different scales

- Methods are available for combining data from scales that are related but have different definitions for their categories (discussed in Anne Whitehead's book – Meta-analysis of controlled clinical trials, section 9.3).

Where next?

- An MRC project.
  - Practical methods for ordinal data meta-analysis in stroke
  - 1 June 2010 to 31 May 2012

a. Review the methods available for meta-analysis of ordinal outcomes.

b. Investigate using each of these methods in real data:
   - how often sufficient data are presented (or can be obtained),
   - how often the available data fulfil any distributional assumptions (and whether there are sufficient data to check assumptions),
   - how easy to understand the results are, and how much detail they show of the way the treatment effect operates.
   - assess the added statistical power gained by using ordinal and continuous data methods over binary methods.

c. Develop a Cochrane workshop on ordinal methods.