Statistically speaking: methodological madness

How to describe your data - the devil is in the distribution

Dear Ms Method Matters,

A recent study published in *Anaesthesia* investigated the effect of anaesthesia before the age of 5 years on the cognition (measured using an intelligence quotient (IQ) test) of children at age 6 [1]. Results showed that mean IQ scores were reduced by 2.1 in children who had been anaesthetised, compared to those who had not. Multivariable regression was also used to look at the effect of premature birth, maternal smoking habits and maternal IQ on children's IQ at 6 years of age.

I have read in several statistical sources that IQ is not a continuous measurement; would it then be correct to present IQ scores as means and standard deviations, and can multivariable regression analyses be applied to non-continuous data? I am also planning a series of experiments on cognitive function in the elderly after anaesthesia and am now totally confused about how the data should be reported for this study.

Coherent Cognition (Sheffield, UK)

Dear Coherent Cognition,

'Levels of measurement' refers to the relationship among the values assigned to the attributes for a variable. The best known classification, with the four levels of measurement (nominal, ordinal, interval and ratio) was developed by psychologist Stanley Smith Stevens in 1946 [2,3]. Intelligence quotient scores fall under ordinal data, as do many of the measurements we might take in the clinic including sedation scores, nausea and vomiting scores, levels of delirium etc. Ordinal data can be viewed as a subset of categorical data, where the variables have natural, ordered categories but the distances between categories are not known and the distance between each category is not the same. For example, in the paper you cited [1], socio-economic data collected from the mothers included their level of education: no schooling; primary school; secondary school; or university. Looking at the categories, there is obviously a logical order to them, as in, ‘secondary school’ would be higher than ‘primary school’, and ‘university’ would be higher than ‘secondary school’, but you cannot say that the difference between ‘no schooling’ and ‘primary school’ is the same as the difference between ‘secondary school’ and ‘university’. That is what we mean when we say, the distance between each category is not the same. When looking at scales such as the Ramsay Sedation Scale (Table 1), it is also obvious that the difference between ‘sluggish response to stimulus’ (score 5) and ‘no response to stimulus’ (score 6) cannot be the same as ‘anxious or restless’ (score 1) and ‘co-operative, oriented and tranquil’ (score 2). For the above examples, data should be expressed as mode (the score with the most frequent occurrence), or percentage frequencies. Another example would be ASA classification. It would be meaningless to express baseline characteristics of your cohort as being mean ASA physical classification 2.74.

This is a contentious topic among statisticians, some of whom believe that ordinal variables should be treated strictly as though they were categorical [4]. But I think that the original data type designations suggested by Smith Stevens may be too restrictive and that there would be many data types which do not fall into any of those pre-specified categories. I believe that not all ordinal data were created equal, and there is a continuum of ordinality. It is obviously not useful to express sedation scores, nor ASA using means and standard deviations, but there are some scales, for example, pain scores, Glasgow Coma Scores and satisfaction scores, which are often expressed using means. There is currently no consensus among statisticians regarding whether this type of can be treated as discrete, and therefore, the mean can be expressed. It is suggested here that the most informative way to express ordinal data, in particular Visual Analogue Scale (for pain) and Glasgow Coma Scores would be to indicate the mode, median and the interquartile range (IQR). The median is the last number in the second quartile, while the interquartile range is the distance between the first and third quartile. The interquartile range for this set of eleven Glasgow Coma scores [3, 8, 8, 9, 10, 11, 12, 12, 13, 13, 15] can be calculated thus; find the median, (in this example, it is 11), find the middle number to the left of the median (here, it is 8), this is your quartile 1 (Q1). Then find the middle number to the right of the median (here, it is 13), this is your quartile 3 (Q3). Subtracting Q1 from Q3 gives you an IQR of 5 [5]. Expressing data as median and IQR is more robust against outliers than expressing mean (standard deviations), and therefore gives a better idea regarding the distribution of your data.

It is common to ask patients their perceived satisfaction regarding hospital services. A patient may be given a questionnaire pre-surgery, and then the same questionnaire following surgery. The response is usually given in the form of a six-point Likert Scale, ranging from 0 (not satisfied) to 5 (very satisfied). As well as expressing the results as the number pf patients who changed their minds negatively, or positively, expressing the results in means pre- and post-surgery would instantly give the reader an idea of whether overall, perceived satisfaction was increased, or decreased post-surgery. What also strengthens the evidence for expressing this ordinal data as though it were continuous is that the test is performed on the same patients, who will probably perceive the scale in the same way each time, making the comparison more valid.

Although more complicated than linear regression, it is possible to conduct ordinal logistic regression, where an ordinal dependent variable can be predicted using interactions between one, or more independent variables [6]. The independent variables could be continuous, nominal, or ordinal. For example, we could predict patient satisfaction score on a six-point scale using type of anaesthetic: local; inhalational; or intravenous. The output table, after performing the ordinal regression function with most statistical software would look similar to Table 2 (please note that figures are simulated). You can see that there are five ‘intercepts’, and these correspond to the five partitions of the six-point satisfaction scale. In this example, patients who underwent surgery with local anaesthesia have been designated the reference category. The effect of inhalational and intravenous anaesthesia on satisfaction scores are then compared to the effect of local anaesthesia. The column of interest in Table 2 is that of Exp(B), which are the results presented as proportional odds ratios. When looking at the significant level for inhalational anaesthesia we can see that the p value is 0.919 (and therefore, is not significant), but the p value for intravenous anaesthesia is 0.001. This indicates that this type of anaesthesia makes a significant contribution to satisfaction scores. We can expect that for a one unit increase in intravenous anaesthesia, i.e., going from 0 (no intravenous anaesthesia) to 1 (intravenous anaesthesia), the odds of the highest satisfaction score versus the combined middle and low satisfaction scores are 3.19 times greater (as seen in the Exp(B) column) compared to local anaesthesia, given all of the other variables in the model are held constant.

Although it is common to see many manuscripts reporting ordinal, or ordinal-like data using means and standard deviation, and that ordinal regression can be performed in a way similar to linear regression, these results should always be interpreted with caution, and we should always keep in mind that the intervals between groups in ordinal variables are not equal.

If you have any questions on methodology, please direct them to Ms Method Matters at *msmethodmatters@gmail.com*

Table 1. Ramsay Sedation Scale

Score	Response
1	Anxious or restless or both
2	Cooperative, orientated and tranquil
3	Responding to commands
4	Brisk response to stimulus
5	Sluggish response to stimulus
6	No response to stimulus

Table 2. Simulated ordinal regression data.

Parameter	B	Std. Error	Hypothesis Test				95% Wald Confidence Interval for Exp (B)	
Threshold			**Wald Chi-square**	**df**	**Sig.**	**Exp (B)**		
Intercept 1	0.151	0.308	0.241	1	0.623	1.660	0.960	2.870
Intercept 2	0.605	0.292	4.310	1	0.038	1.163	0.636	2.128
Intercept 3	-0.323	0.082	15.605	1	0.000	1.832	1.034	3.245
Intercept 4	-0.132	0.028	22.651	1	0.000	0.724	0.617	0.850
Intercept 5	1.798	1.575	1.303	1	0.254	0.876	0.830	0.925
Local	0	.	.	.	.	1	.	.
Inhalational	0.037	0.3636	0.010	1	0.919	1.038	0.509	2.116
Intravenous	1.161	0.3446	11.358	1	0.001	3.194	1.626	6.277

References

1. de Heer IJ, Tiemeier H, Hoeks SE, Weber F. Intelligence quotient scores at the age of 6 years in children anaesthetised before the age of 5 years. *Anaesthesia* 2017; **72:** 57-62.
2. . Level of MeasurementLevel of measurement. In: Kirch W, ed. *Encyclopedia of Public Health.* Dordrecht: Springer Netherlands, 2008: 851-2.
3. Stevens SS. On the Theory of Scales of Measurement. *Science* 1946; **103:** 677-80.
4. Jakobsson U. Statistical presentation and analysis of ordinal data in nursing research. *Scandinavian Journal of Caring Science* 2004; **18:** 437-40.
5. McCluskey A, Lalkhen AG. Statistics II: Central tendency and spread of data. *Continuing Education in Anaesthesia Critical Care and Pain* 2007; **7:** 127-30.
6. Norris CM, Ghali WA, Saunders LD, et al. Ordinal regression model and the linear regression model were superior to the logistic regression models. *Journal of Clinical Epidemiology* 2006; **59:** 448-56.

Acknowledgements

SC is statistical advisor to Anaesthesia. No other competing interests declared.

S. W. Choi

Assistant Research Officer

Department of Anaesthesiology,

The University of Hong Kong,

Hong Kong, HKSAR

Email: htswchoi@hku.hk

D. M. H. Lam

Resident

Department of Anaesthesiology,

Queen Mary Hospital,

Hong Kong, HKSAR