Statistics in Medicine

Ordinal scale and statistics in medical research

MISSER FORREST, BJØRN ANDERSEN

Abstract

A survey of the 1982 editions of 12 medical journals showed that in at least 70% of 175 papers employing ordinal measurement scales, statistical methods were used, which do, in fact, assume a more refined measurement scale. Non-parametric methods suited for analysis of ordinal data are listed.

Introduction

Methodological errors are abundant in medical research.[1] Prominent among them is inappropriate use of statistical methods.[2-6] One potential error has received little attention: the use of statistical methods that assume too high a level of measurement. Its importance is apparent as textbooks classify tests according to the level of measurement.[7 8]

The concept of relations between measurement scales and statistical methods is attributed to Stevens,[9] who identified four scales of ascending level. In the nominal scale observations are in mutually exclusive and exhaustive classes. Order, interval, and ratio are meaningless. Medical examples are diagnoses and treatments. The ordinal or ranking scale also has mutually exclusive and exhaustive classes, but there is order between them, as indicated by expressions such as "higher," "better," or "more." Medical examples are Dukes' classification and grading of hypertensive retinopathy. More refined measurements are the interval and ratio scales. They are continuous and have a measurement unit such as: cm, g, sec, °C, and derivatives thereof. Body temperatures in °C are interval (the zero point is arbitrary) and haemoglobin concentrations are ratio scale measurements. On interval scales the differences between values have the same meaning whatever the size of the measurements, whereas on ratio scales the ratio of the two values has a constant meaning.

A recent survey of research articles in volume 306 of the New England Journal of Medicine showed that a fifth of articles contained data on an ordinal scale.[10] In no instance were analyses performed with statistical methods specifically meant to deal with ordinal data. Some authors collapsed categories to a nominal scale with two levels to apply contingency table analyses, while others used methods such as Student's t test, which assume ratio or interval scales. The nominal scale may entail a loss of information, while the ratio and interval scales introduce untenable assumptions.

We surveyed published papers from several journals covering various aspects of medicine to determine which statistical methods were currently used for descriptive as well as analytical purposes, whenever measurements were on an ordinal scale.

Department of Surgical Gastroenterology D, Herlev Hospital, University of Copenhagen, DK-2730 Herlev, Denmark
MISSER FORREST, MD, research fellow
BJØRN ANDERSEN, MD, PHD, chief surgeon

Correspondence to: Dr Forrest.

Materials and methods

We surveyed all 1982 volumes of the following 12 medical journals (and numbers of original reports): New England Journal of Medicine (152), Annals of the Rheumatic Diseases (113), American Journal of Obstetrics and Gynecology (425), Acta Medica Scandinavica (136), British Journal of Surgery (232), Acta Anaesthesiologica Scandinavica (120), British Medical Journal (551), Lancet (287), Clinical Pharmacology and Therapeutics (203), British Journal of Clinical Pharmacology (169), British Journal of Dermatology (159), and Annals of Internal Medicine (141). The choice was arbitrary, being principally determined by ease of access, but it covered both specialised publications and those addressed generally to the medical profession. Among 2688 original reports we selected those 175 research reports in which at least one criterion was measured on an ordinal scale. The statistical methods used either to describe or to analyse those data were registered. Whenever several criteria were combined to form a total score, we treated the resulting data as on an ordinal scale, provided at least one of the individual items was on an ordinal scale.

Results

We found five different types of ordinal scales: (a) the established scales such as Dukes', retinopathy, Visick grading, Ritchie index, Karnofsky rating, and Killip classes; (b) the combined rating scales such as the Hamilton scale; (c) the transformed ratio and interval scales that occurred whenever ratio and interval data were represented by classes of unknown or unequal spans—for example, young, middle aged, elderly, and old; (d) the ad hoc scales describing phenomena by "absent," "slight," "moderate," and so on; and (e) the visual analogue scale. Although usually converted to either cm or per cent the visual analogue scale has no true unit of measurement and is accordingly ordinal only. Distribution of studies according to this division was: established scales 23, combined scales 31, transformed ratio and interval scales 7, ad hoc scales 91, and visual analogue scales 41, totalling 193 in 175 papers because some papers contained more than one type of scale.

We listed statistical methods used for descriptive and analytical purposes. Central tendency descriptive statistical methods were: mean (69 reports), median (10), percentage (22)—for example, 85% of patients were cured or at least noticeably improved—and others (three). Dispersion descriptive statistical methods were: SD (37 papers), SEM (23) used evidently for descriptive purposes only, SD or SEM (seven)—for example, $7·4\pm3·1$ without translation to either mean SD or mean SEM—range (11), and quantiles (six). Parametric analytical statistical methods were: t test (54 reports), ANOVA (26), ANCOVA (seven), correlation (11), regression (four), factor analysis (two), confidence limits (six), and others (10). Non-parametric analytical statistical methods were: χ^2 and Fisher's exact test (34 reports), Mann-Whitney and Wilcoxon (52), Kruskal-Wallis and Friedman (16), correlation (seven), regression (two), and others (five). While 28 (16%) of the 175 papers were without any statistical methods, 70 (40%) used parametric methods, 51 (29%) non-parametric methods, and 26 (15%) a mixture. The figure gives the methods we use in ordinal data.
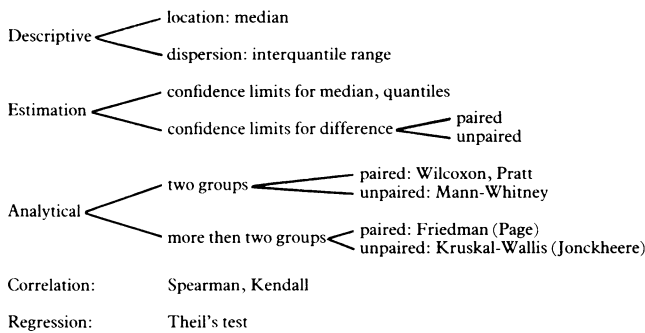
Discussion

Ordinal scales are usually necessary because of the absence of any precise method of measurement, but the choice of descriptive measure, such as the mean, SD, and median, depends on the scale of measurement.[3 11] Traditionally, the classes of an ordinal scale are represented by the numerals 1,2,3, and so on, but we cannot

properly believe that the scale now possesses the properties of ratio and interval scales. With ordinal data it is likewise incorrect to join points on a graph—for example, if showing the mean height by social class. Particular problems exist with combined rating scales.

Descriptive — location: median
          — dispersion: interquantile range

Estimation — confidence limits for median, quantiles
           — confidence limits for difference — paired
                                                — unpaired

Analytical — two groups — paired: Wilcoxon, Pratt
                          — unpaired: Mann-Whitney
           — more then two groups — paired: Friedman (Page)
                                    — unpaired: Kruskal-Wallis (Jonckheere)

Correlation: Spearman, Kendall

Regression: Theil's test

Some simple methods useful in analysis of ordinal data.[16]

When several items are measured on ordinal scales it is far from certain that the sum of scores has even ordinal properties. If items are correlated or have different weights reductions may appear in the total score without any simple and clear cut relation to fluctuations in the patient's overall condition. Numbers with several digits may, in fact, represent nothing more than a nominal scale—for example, licence plates on vehicles. Ordinal scales may represent elastic "rulers." Any numerical translation retaining order will, a priori, be as "true" as any other. Conclusions based on such translations may, however, vary: suppose four patients were admitted to hospital simultaneously and discharged after treatment in the sequence A, B, C, and D. As A had a shorter stay than B, who had shorter stay than C, who had shorter stay than D measurements are ordinal. A and B had shorter stays than C and D, but which of the pairs A and D and B and C had the longer stay? If A stayed say one day, B three, and C four opposing conclusions will be reached depending on whether D stayed for more or less than six days.

Consequently, descriptive statistical measures such as mean and SD, which require addition of observations, are invalid whenever data are on ordinal scales. It seems evident that methods which, like the t test, analysis of variance, and regression, are based on means and SDs would also be invalid. This contention is supported by the fact that those methods assume a gaussian distribution of data, which is meaningless without measurement units on the abscissa. The question is, however, disputed among statisticians.[12 13]

Moses et al stated that parametric methods such as the t test may serve as an approximation (of the Mann-Whitney test).[10] As the numerical translation of ordered categories is arbitrary, different analyses of a set of ordered data may generate different p values and thus lead to different conclusions. Lord calculated mean and standard deviation from the numbers football players wear on their uniform.[14] Faced with the objection that cardinal numbers (nominal scale) could be neither added nor multiplied he replied: "Since the numbers do not remember where they came from, they always behave just the same way, regardless." Baker, Hardyck, and Petrinovich used Monte Carlo methods to evaluate the performance of the t test in 35 different transformations to ordinal scales of measurements that were ratio and interval scales and confined to a normal, rectangular, or exponential distribution.[15] The essential conclusion was that the t test was reasonably reliable in the circumstances examined. This was indeed reassuring, although findings need not pertain to parametric methods other than the t test or other circumstances.

Since ordinal data are often used in clinical research further studies on the behaviour of both parametric and non-parametric statistical methods should be done whenever truly continuous phenomena such as pain or mood are measured on an ordinal scale. It was, however, reassuring to find that non-parametric methods were often used in the statistical analyses of medical research.

References

1 Schor S, Karten I. Statistical evaluation of medical journal manuscripts. JAMA 1966;195:145-50.
2 Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in BMJ from January to March 1976. Br Med J 1977;i:85-7.
3 White S. Statistical errors in papers in the British Journal of Psychiatry. Br J Psychiatry 1979;135:336-42.
4 Glantz SA. Biostatistics: how to detect, correct and prevent errors in the medical literature. Circulation 1980;61:1-7.
5 Altman DG. Statistics in medical journals. Statistics in Medicine 1982;1:59-71.
6 McArthur RD, Jackson GG. An evaluation of the use of statistical methodology in the Journal of Infectious Diseases. J Infect Dis 1984;149:349-54.
7 Siegel S. Nonparametric statistics for the behavioral sciences. New York: McGraw Hill, 1956.
8 Sokal RR, Rohlf FJ Biometry. 2nd ed. New York: W H Freeman and Company, 1981.
9 Stevens SS. On the theory of scales of measurement. Science 1946;103:677-80.
10 Moses LE, Emerson JD, Hosseini H. Analyzing data from ordered categories. N Engl J Med 1984;311:442-8.
11 Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. Br Med J 1983;286:1489-93.
12 Gaito J. Measurement scales and statistics: resurgence of an old misconception. Psychol Bull 1980;8:564-7.
13 Labovitz S. Statistical usage in sociology. Sociol Methods Research 1972;1:13-37.
14 Lord FM. On the statistical treatment of football numbers. Am J Psychol 1953;8:750-1.
15 Baker BO, Hardyck CD, Petrinovich LF. Weak measurements vs. strong statistics: an empirical critique of S S Stevens' proscriptions on statistics. Educational and Psychological Measurement 1966;26:291-309.
16 Bradley JA. Distribution-free statistical tests. New Jersey: Prentice-Hall Inc, 1968.

How effective is the homoeopathic pertussis vaccine and are there any side effects?

Pertussin is a homoeopathic "nosode"—that is, sputum from a patient suffering from pertussis that has been subjected to the homoeopathic method of preparing medicine. It is taken by mouth. Usually it is given to patients suffering from the sequelae of whooping cough but some practitioners have given it in an attempt to prevent children from contracting the disease. A common dosage schedule is to give Pertussin 30C, three doses at 12 hour intervals, followed by one dose every month for the next four months. In common with other homoeopathic medications Pertussin is safe and free from side effects, but its efficacy as a preventive is uncertain as no clinical trials have been carried out. There is no reason why this form of homoeopathic immunisation should not be given to children who are considered unsuitable for the orthodox vaccine or whose parents refuse orthodox immunisation, but it is essential in such cases to explain frankly the lack of any evidence for the efficacy of the vaccine.—ANTHONY CAMPBELL, consultant physician, London.

Do antibiotic drugs reduce the effectiveness of oral contraceptives?

It is not possible to give a definite answer to this question. There have been sporadic reports of women taking oral contraceptives who have become

pregnant when given antibiotics; but considering the number of women who could be at risk from this interaction, the number of reports is small. The Committee on Safety of Medicines has had more than 40 reports of pregnancy attributed to antibiotics prescribed for women taking the pill, but a causal relation cannot be assumed from these reports. The antibiotics most often implicated are broad spectrum oral penicillins, tetracyclines, and co-trimoxazole. There is a possible mechanism for this interaction. Ethinyl oestradiol undergoes enterohepatic recirculation, a process that requires bacterial deconjugation of the conjugated drug in the intestine. In animals there is evidence that antibiotics reduce the enterohepatic recirculation of ethinyl oestradiol. This probably occurs also in man, but it is doubtful whether the effect on circulating ethinyl oestradiol concentration would be of sufficient magnitude to interfere with contraceptive efficacy—at least in most women. Studies with ampicillin in women have not shown any significant interaction with oral contraceptives,[2] but this does not exclude the possibility that a few women may be at risk. Progestogens do not undergo enterohepatic recirculation as unchanged drugs and progestogen only contraceptives should, therefore, not be affected by antibiotics.—LINDA BEELEY, consultant clinical pharmacologist, Birmingham.

1 Orme M L'E, Back DJ, Breckenridge AM. Clinical pharmacokinetics of oral contraceptive steroids. Clin Pharmacokinet 1983;8:95-136.
2 Back DJ, Breckenridge AM, MacIver M, et al. The effects of ampicillin on oral contraceptive steroids in women. Br J Clin Pharmacol 1982;14:43-8.