

death follows the withdrawal of ventilatory support, however, you could argue that death must be the intended result and not merely a side effect of acting in the patient's best interests. This is because paralysing drugs suppress a patient's breathing for some time after an infusion has been stopped. Because the patient completely depends on the ventilator she dies quickly when it is withdrawn. The BMA suggests that withdrawing respiratory support in these circumstances could be interpreted in law as an intended killing.³

The doctrine of double effect is difficult to apply in this case, not impossible. To avoid confusion it is necessary to deal with each decision in turn. The first decision—to use muscle relaxants—itself has dual effects, the intended one being to make ventilation possible, and the side effect being to make the patient completely dependent on it. The second decision—to withdraw ventilation—has the effect that the patient dies quickly, but this doesn't have to be the intended effect. The two decisions are morally independent, despite being chronologically and logically sequential, because the intention behind the first decision (to make ventilation possible) is not linked to the side effect of the second, the patient dies. This could be an accurate description of the doctor's intentions in the face of difficult circumstances. It could also provide a legal defence against the charge of murder, although I know of no case law testing this interpretation.

The alternative would be either to continue respiratory support indefinitely or to stop the paralysing drugs first and wait for any effects to wear off completely before withdrawing ventilation. The first is arguably not in the patient's best interests. The other may prolong the patient's suffering and that of her parents because of the delay between deciding to withdraw life prolonging treatment and actually doing it. The only reason why muscle relaxants were used in the first place was to allow ventilatory support, so the intention behind withdrawing it can only be to allow the subsequent withdrawal of ventilation, thereby linking the intention of the first decision with the effect of the second. The patient may become distressed and, without the effect of the muscle relaxants, the ventilator may not work properly anyway. It is difficult to see how withdrawing muscle relaxants can be in a patient's best interests without taking him or her off the ventilator at the same time.

In any case there are drugs, such as atracurium, that do not have these problematic residual effects, making this particular moral difficulty disappear.

Competing interests: None declared.

- 1 Zucker MB, Zucker HD. *Medical futility and the evaluation of life-sustaining interventions*. Cambridge: Cambridge University Press, 1997.
- 2 R v Woollin [1998] 4 All ER 103.
- 3 Section 17.A. *Withholding and withdrawing life-prolonging medical treatment: guidance for decision making*. London: BMJ Books, 1999.

Mann-Whitney test is not just a test of medians: differences in spread can be important

Anna Hart

The Mann-Whitney (or Wilcoxon-Mann-Whitney) test is sometimes used for comparing the efficacy of two treatments in clinical trials. It is often presented as an alternative to a *t* test when the data are not normally distributed. Whereas a *t* test is a test of population means, the Mann-Whitney test is commonly regarded as a test of population medians. This is not strictly true, and treating it as such can lead to inadequate analysis of data.

Use of Mann-Whitney test

The Mann-Whitney test is a test of both location and shape. Given two independent samples, it tests whether one variable tends to have values higher than the other. As Altman states, one form of the test statistic is an estimate of the probability that one variable is less than the other,¹ although this statistic is not output by many statistical packages. In the case where the only distributional difference is a shift in location, this can indeed be described as a difference in medians. Hence, for example, the online help facility in Minitab 10.51 states that the Mann-Whitney test is “a two-sample rank test for the difference between two population medians ... It assumes that the data are independent random samples from two populations that have the same shape.” Figure 1 shows two distributions for which this

Summary points

The Mann-Whitney test is used as an alternative to a *t* test when the data are not normally distributed

The test can detect differences in shape and spread as well as just differences in medians

Differences in population medians are often accompanied by equally important differences in shape

Researchers should describe the clinically important features of data and not just quote a P value

is the case. One distribution is shifted 0.75 units to the right: the medians differ by 0.75 units but the shapes are identical.

Theoretically, in large samples the Mann-Whitney test can detect differences in spread even when the medians are very similar. However, an alternative form of the test is better than the standard Mann-Whitney test for this purpose.² The alternative test, however, is not

Statistics Group,
Faculty of Science,
University of
Central Lancashire,
Preston PR1 2HE
Anna Hart
principal lecturer
ahart@uclan.ac.uk

BMJ 2001;323:391-3

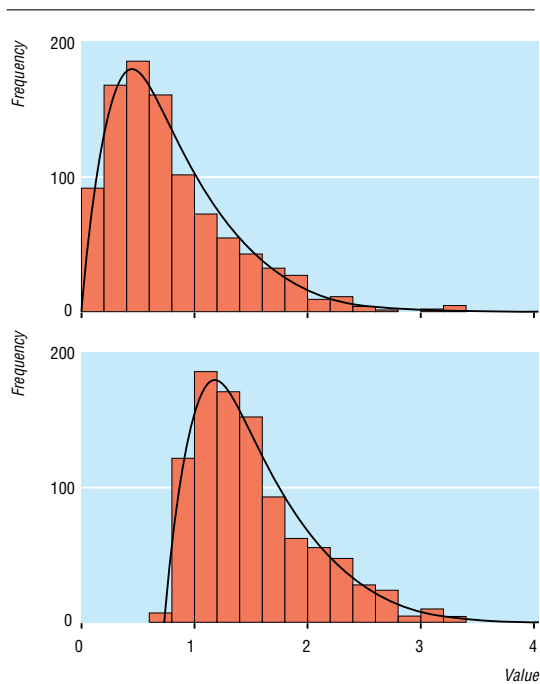


Fig 1 Two distributions with a difference in median but no difference in shape and spread

very efficient when population medians are unequal and is not widely available in statistical packages.

Differences in population medians are often accompanied by other differences in spread and shape. Moreover, the difference in medians may not be the most striking or the most clinically important difference. It is important to look at distributional differences and discuss them. Figure 2 shows an example in which the median values are 0.65 and 1.14 units. The distribution with the larger median also has larger spread. The spread is shown clearly in figure 3, which shows box plots of samples of 25 drawn from these two distributions. (The P value from the Mann-Whitney test is 0.02.) If the difference is assumed to be merely a difference in medians other clinically important information could be ignored.

Methods

I examined the use of the Mann-Whitney test in papers published in the *BMJ* between September 1999 and August 2000. I did an online search of the electronic text of the journal using the keywords Wilcoxon, Mann, and Whitney. I identified five papers that had used the Mann-Whitney test but where, in my judgment, the information given suggested that there might be important distributional differences other than a shift in location. These are described briefly below.

Examples

Grande et al studied the impact on place of death of a hospital at home service for palliative care.³ The authors noted a significant difference among patients randomised to hospital at home care: "Patients in the hospital at home group who were admitted to the service survived significantly longer after referral than

hospital at home patients who were not admitted (16 v 8 days)." There were 112 patients admitted to the service (median survival 16 days, interquartile range 5-42.5) and 73 patients who were not admitted (8, 3-18 days). The striking feature about these three sets of summary statistics is that each in the former group is about twice that for the second group. This suggests that the difference between the two distributions might not be just a shift of 8 days: the difference might be multiplicative, not additive—that is, patients who were admitted might survive twice as long as those who were not admitted.

Williams et al did a cost effectiveness study of open access follow up for inflammatory bowel disease.⁴ One of the measures was the total cost of secondary care, and this was compared for two groups: open access and routine visit. The mean (SD) cost was £582 (£807.94) for the 77 patients in the open access group and £611 (£475.47) for the 78 patients in the routine visit group. Although the mean is higher in the second group, the standard deviation is much higher in the first. There must, therefore, have been some very large values in the first group. Without further information it is difficult to be sure, but there seem to be distributional differences between the two groups. The choice of a Mann-Whitney test for these economic data has been criticised elsewhere.⁵ If total expenditure is the aspect of prime interest then a *t* test would have been more appropriate.⁶ If the interest lay in the distributions, it is unlikely that the medians alone would adequately have described the differences.

Lux et al studied responses of local research ethics committees.⁷ A conclusion was that "The required number of complete copies of protocols and documents ... was significantly lower for the local committees that used a fast track system." The 44 committees in the fast track group required a median of three copies (95% percentiles 2 and 13) compared

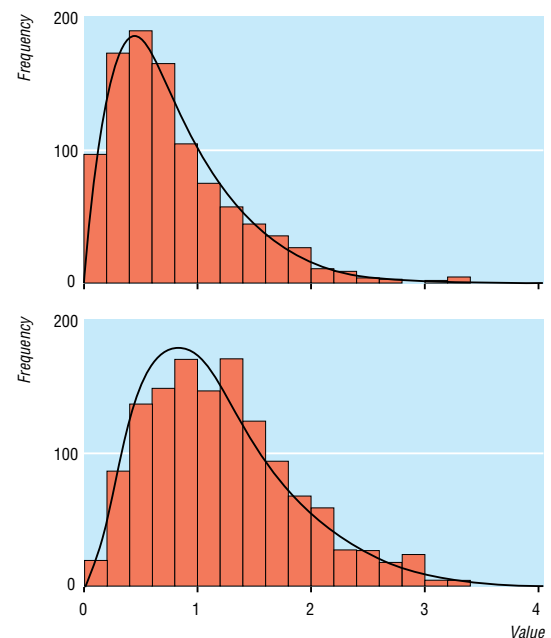


Fig 2 Two distributions with different medians and different shapes. The distribution with the larger median also has a greater spread

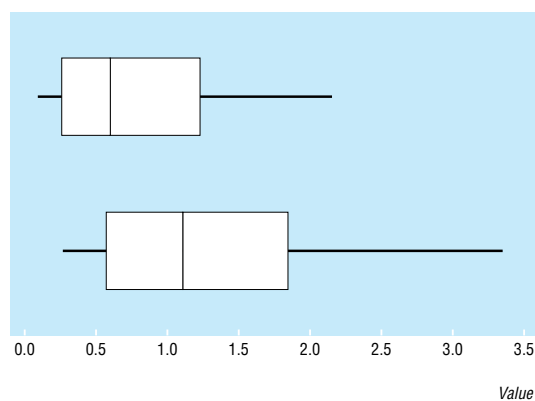


Fig 3 Box plots of samples of size 25 drawn from the distributions in figure 2. Vertical lines indicate the medians and boxes the interquartile range

with 11 (1 and 15) copies for the 55 committees in the standard group. Not only are the medians different, the distributions must also be different. About half of the fast track committees asked for two or three copies, whereas about half of the other committees asked for 11-15 copies. These differences, which the authors did not comment on, relate to shape as well as location of the distributions.

Macleod et al studied women with breast cancer from affluent and deprived areas.⁸ One of their conclusions is “The time between the date of the referral letter and the first clinic was one day shorter in women from affluent areas.” The median (interquartile range) time was 6 (1-13) days in the affluent area and 7 (4-20) days in the deprived area. Although the medians differ by one day, the summary statistics suggest that the data for the deprived group are more right skewed, and differences between the two groups might be much more pronounced for the higher waiting times. It would have been helpful to discuss this in the paper.

A similar feature is even more evident in data from a study of pain in blood glucose testing.⁹ A visual analogue scale was used to record pain at the ear or thumb. The authors report “The median pain score was 2 mm in the ear group and 8.5 mm in the thumb

group ... the difference in median pain score is small.” Although this is true, the box plots in the paper show that the spread of scores in the thumb group is much greater than for the ear group. In particular, at least three out of 30 people in the thumb group report a score that is at least twice the highest value in the ear group. Overall, values seem much higher in the thumb group. This is important because patients are likely to be more concerned with the worst pain they might experience than the median value.

Recommendations

Researchers should take care to describe their data and to be clear about the features that are most clinically important. They should use the statistical test that is most relevant for their hypotheses, and describe the features of the data that are likely to have caused a hypothesis to be rejected. As is always the case, it is not sufficient merely to report a P value. In the case of the Mann-Whitney test, differences in spread may sometimes be as clinically important as differences in medians, and these need to be made clear to the reader.

Funding: None.

Competing interests: None declared.

- 1 Altman DGA. *Practical statistics for medical research*. London: Chapman and Hall, 1991.
- 2 Sprent P. *Data driven statistical methods*. London: Chapman and Hall, 1998.
- 3 Grande GE, Todd CJ, Barclay SIG, Farquhar MC. Does hospital at home for palliative care facilitate death at home? Randomised controlled trial. *BMJ* 1999;319:1472-5.
- 4 Williams JG, Cheung WY, Russell IT, Cohen DR, Longo M, Lervy B. Open access follow up for inflammatory bowel disease: pragmatic randomised trial and cost effectiveness study. *BMJ* 2000;320:544-8.
- 5 Barber JA, Thompson SG. Open access follow up for inflammatory bowel disease. Would have been better to use *t* test than Mann-Whitney U test. *BMJ* 2000;320:1730.
- 6 Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? *BMJ* 2000;320:1197-200.
- 7 Lux AL, Edwards SW, Osborne JP. Responses of local research ethics committees to a study with approval from a multicentre research ethics committee. *BMJ* 2000;320:1182-3.
- 8 Macleod U, Ross S, Twelves C, George WD, Gillis C, Watt GC. Primary and secondary care management of women with early breast cancer from affluent and deprived areas: retrospective review of hospital and general practice records. *BMJ* 2000;320:1442-5.
- 9 Carley SD, Libetta C, Flavin B, Butler J, Tong J, Sammy I. An open prospective randomised trial to reduce the pain of blood glucose testing: ear versus thumb. *BMJ* 2000;321:20.

(Accepted 20 February 2000)

When I use a word ...

Attendee

Well, I don't. The verb to attend comes from the Latin, meaning “to stretch to.” From this developed the meaning “to direct the mind or energy to,” and later “to direct one's care to.” The meaning “to present oneself at a meeting” is dated back to the 17th century.

The suffix -ee is given four meanings in the *Shorter Oxford Dictionary* and is admittedly a muddy area. However, none of the meanings comes near to denoting the subject of a verb (the person who does the thing, such as attending a meeting). It is I who attends the meeting, and not the meeting which attends me. The nearest legitimate use listed would be in the adoption into English of the past participle of certain reflexive verbs in French, such as “se refugier” (to remove oneself from a place in the interests of one's safety), which results in the word refugee. While I have taken refuge from meetings at times, I have never thought of attending a meeting as a reflexive activity. Perhaps “attendee”

should be reserved for those who attend meetings in order to listen to themselves.

I may have occasionally been the smallest person at a meeting and might therefore qualify for the diminutive suffix, rather as a small coat becomes a coatee, but, since I was never an attend, even being small would not make me an attendee.

There are already two words for a person who attends, and they are attendant and attender. Curiously the *Shorter Oxford Dictionary* gives the former the meaning at issue, whereas I find the second the easier to use. But surely there is no need for a third.

My suggestion is that we learn to accept, when talking about what we do, that we probably haven't invented a new practice and so we probably don't need a new word.

Andrew West *specialist registrar in child and adolescent psychiatry, Oxford*