# Comparing the effects of two treatments on two ordinal outcome variables

Horstmann, Vibeke; Lanke, Jan

2015

Working Papers in Statistics
No 2015:16

Department of Statistics
School of Economics and Management
Lund University

# Comparing the effects of two treatments on two ordinal outcome variables

VIBEKE HORSTMANN, LUND UNIVERSITY

JAN LANKE, LUND UNIVERSITY

# Comparing the effects of two treatments on two ordinal outcome variables

**Vibeke Horstmann and Jan Lanke**[1]

Department of Statistics, Lund University. Box 743, SE-220 07 Lund, Sweden

**Running title: Two treatments, two ordinal variables**

*Key words:* Changes; Comparisons; Effects; Ordinal;

## 1 Introduction

When evaluating whether the effect of one treatment is larger than that of another the first step in the comparison is to decide what should be understood by the statement that one patient has achieved a greater effect than has another patient. When the outcome variable is quantitative, measured on a ratio scale, absolute or relative effects are the most commonly used effect measures; however, such effects are usually not meaningful for ordinal outcome variables. In order to answer the question whether one of two treatments acts more effectively on one of two outcome variables and the other treatment more efficiently on the other we shall present a method of comparing the treatment effects of patients that is based on pair-wise comparisons between patients in analogy with many non-parametrical methods. These comparisons use only the ordinal properties of the outcome variables. We shall even define a measure of the difference between the treatment effects and demonstrate how confidence intervals can be constructed.

## 2 Two patients, two time points, just one outcome variable

Although our aim is to relate changes in one variable, $x$, to those in another, $y$, we shall in this section consider just one single variable, say $x$.

Since the magnitude of a change in an ordinal variable is usually undefined, special care is necessary when we want to give the verdict ″one drop is larger than another″.

---

[1]Corresponding author: Jan.Lanke@stat.lu.se

Let us consider two individuals, $pat_1$ and $pat_2$, and two time points, $t'$ and $t''$ ($>t'$), giving rise to the set of observations

$$\{x_1', x_1'', x_2', x_2''\}.$$

The changes from one time point to the other will be considered in terms of drops. With definitions given below we shall use the notation

$$x_1 \, D \, x_2$$

(D for "drop") to indicate that the drop in the $x$-variable from $t'$ to $t''$ is greater for $pat_1$ than for $pat_2$, and

$$x_1 \, d \, x_2$$

to indicate that the drop is smaller for $pat_1$, or, equivalently, that the drop is greater for $pat_2$. Furthermore

$$x_1 \, E \, x_2$$

(E for "equal") indicates that the change is the same for the two patients. Finally for all remaining situations, we use the notation

$$x_1 \, N \, x_2$$

(N for "nothing").

The case $x_1 \, D \, x_2$ will be said to occur if either

(Dt) $\quad x_1' \geq x_2' \;\&\; x_1'' \leq x_2''$ with at least one strict inequality

(t for "within timepoint") or

(Dp) $\quad x_1' \geq x_1'' \;\&\; x_2' \leq x_2''$ with at least one strict inequality

(p for "within patient"). (Dt) will thus occur if at $t'$ the value of $x$ is larger for $pat_1$ than for $pat_2$ but at $t''$ the reverse inequality holds. (Dp) occurs when the changes in x are in opposite directions for the two individuals and $pat_1$ is the one who experiences a decrease. Note that (Dp) and (Dt) can occur at the same time.

Analogously, cases $x_1 \, d \, x_2$, where $pat_1$ has a smaller decrease in $x$ than $pat_2$, are said to occur when

(dt) $x_1' \le x_2'$ & $x_1'' \ge x_2''$ with at least one strict inequality

or

(dp) $x_1' \le x_1''$ & $x_2' \ge x_2''$ with at least one strict inequality.

Again, (dt) and (dp) can occur at the same time.

Finally, we use $x_1$ E $x_2$ to denote that either

(Et) $x_1' = x_2'$ & $x_1'' = x_2''$

or

(Ep) $x_1' = x_1''$ & $x_2' = x_2''$.

Thus (Et) happens when the patients have identical values at start and at also the end while (Ep) happens when there is no change in the *x*-variable for any of the patients.

The remaining cases, denoted by $x_1$ N $x_2$, are those where we do not find it possible to present a convincing argument why one of the two decreases should be considered larger than the other, neither do we want to consider the changes identical.

Let us consider e.g. the relation D more closely. If $x_1$ D $x_2$ then the drop in $x_1$ is greater than that in $x_2$ in a very strong sense indeed. In particular it is easy to check that if *x* happens to be a ratio-scale variable then, whenever we have $x_1$ D $x_2$, it also holds that $x_1$ has had a greater drop than $x_2$ both in the absolute and relative sense. On the other hand, one can find examples showing that the fact that $x_1$ has had a greater drop than $x_2$, be it absolute or relative or even both, does not imply that $x_1$ D $x_2$. Also the relation $x_1$ E $x_2$ implies that $x_1$ and $x_2$, considered as variables on an appropriate scale, have shown the same change, whether in the absolute or relative sense.

We have defined four relations between two patients:

$x_1$ D $x_2$ ; $x_1$ d $x_2$ ; $x_1$ E $x_2$ ; $x_1$ N $x_2$.

From the definition of the last of these four, it is clear that always at least one of them holds. Now we want to show that always exactly one of them holds; it is then sufficient to demonstrate that it can never happen that two of them are fulfilled simultaneously. Again, it is clear that N is incompatible with each of D, d and E.

What remains to be shown thus is that no two of D, d and E can happen at the same time.

First, suppose that $x_1$ D $x_2$ occurs and that it is (Dt) that holds. Obviously, neither (dt) nor (Et) can then occur; it remains to verify the same for (dp) and (Ep). If both (Dt) and (dp) hold, then

$$x_1' \geq x_2' \;\&\; x_1'' \leq x_2'' \quad \text{and} \quad x_1' \leq x_1'' \;\&\; x_2' \geq x_2''$$

and hence

$$x_1' \leq x_1'' \leq x_2'' \leq x_2' \leq x_1' \,,$$

that is all four values must be equal, which contradicts both (Dt) and (dp). In the same way, the joint occurrence of (Dt) and (Ep) would give a contradiction.

If instead it was (Dp) that holds then trivially neither (dp) nor (Ep) could hold and, in the same way as above, both (dt) and (Et) can be shown to be impossible.

Just by reversing the roles of $x_1$ and $x_2$ one can see that (d) is incompatible not only with (D), which we already know, but also with (E), and hence we have completed the proof that always exactly one of D, d, E and N holds.

The first two of our four relations,

$$x_1 \,\mathrm{D}\, x_2 \;;\quad x_1 \,\mathrm{d}\, x_2 \;;\quad x_1 \,\mathrm{E}\, x_2 \;;\quad x_1 \,\mathrm{N}\, x_2 \,,$$

when described verbally, give the impression of being order relations while the third has the flavour of an equivalence relation. These impressions are correct; formal proofs of these statements are straightforward, although very tedious.

## 3 Two patients, two time points, two outcome variables

Although our final aim is to compare two groups of persons in terms of their changes in two variables, we shall, in this section, consider just two persons, still called $\mathrm{pat}_1$ and $\mathrm{pat}_2$. What is new is that now we have two variables, $x$ and $y$.

Of course it would be absurd to compare one value of $x$ with one of $y$: the two variables may well relate to quite different types of quantity. For the same reason, a

within-person comparison between the change in *x* and that in *y* would be meaningless. What we want is to identify cases where, compared with $pat_2$, $pat_1$ has a greater drop in one variable but a smaller drop in the other.

For one variable, *x*, we have classified the relation between the changes for two individuals, $pat_1$ and $pat_2$, into four categories

$$x_1 \text{ D } x_2\,, \quad x_1 \text{ d } x_2\,, \quad x_1 \text{ E } x_2 \text{ and } \quad x_1 \text{ N } x_2.$$

For a second outcome variable, y, we similarly have

$$y_1 \text{ D } y_2\,, \quad y_1 \text{ d } y_2, \quad y_1 \text{ E } y_2 \text{ and } \quad y_1 \text{ N } y_2.$$

If $x_1$ D $x_2$ and $y_1$ D $y_2$ happen at the same time, the conclusion to be drawn is that $pat_1$ drops more than $pat_2$ in both *x* and *y***;** this clearly tells us something about $pat_1$ *versus* $pat_2$ but nothing of interest concerning *x versus y*. Similarly the combination $x_1$ d $x_2$ & $y_1$ d $y_2$ is uninformative about the relation of *x* to *y*, and so is $x_1$ E $x_2$ & $y_1$ E $y_2$.

However, for a situation with $x_1$ D $x_2$ & $y_1$ d $y_2$ the conclusion is that, compared with $pat_2$, $pat_1$ has a larger drop in *x* and a smaller drop in *y*, a statement that concerns the changes of the outcomes *x* and *y*. This situation is thus an example of those cases we wanted to identify. The situation $x_1$ D $x_2$ & $y_1$ E $y_2$ shows that $pat_1$ has a larger drop in *x* and an equal drop in *y*, and when $x_1$ E $x_2$ & $y_1$ d $y_2$ holds, $pat_1$ has the same drop in *x* and a smaller in *y*. These last two situations lead to the conclusion that the drop in *x* for $pat_1$ is at least not smaller than that for $pat_2$, while the drop in *y* is at least not larger than the drop for $pat_2$, and the two drops are not equal. Thus, the difference in the effects on *x* and *y* is perhaps slightly less convincingly established in the last two cases.

For the situations $x_1$ d $x_2$ & $y_1$ D $y_2$, $x_1$ E $x_2$ & $y_1$ D $y_2$, and $x_1$ d $x_2$ & $y_1$ E $y_2$ there are equivalent conclusions to be drawn. Finally, the cases with either $x_1$ N $x_2$ or $y_1$ N $y_2$ will be inconclusive.

All the 16 cases are described below; as is seen, six of them tell us something that is of interest in the present context, two of them ($x_1$ D $x_2$ & $y_1$ d $y_2$ ; $x_1$ d $x_2$ & $y_1$ D $y_2$ ) perhaps more strongly than the other four.

Drop for $pat_1$, relative to $pat_2$:

| | $y_1$ D $y_2$ | $y_1$ d $y_2$ | $y_1$ E $y_2$ | $y_1$ N $y_2$ |
|---|---|---|---|---|
| $x_1$ D $x_2$ | ———— | **larger *x*-drop, smaller *y* -drop** | **larger *x*-drop, equal *y* -drop** | ———— |
| $x_1$ d $x_2$ | **smaller *x*-drop, larger *y*-drop** | ———— | **smaller *x*-drop, equal *y* -drop** | ———— |
| $x_1$ E $x_2$ | **equal *x*-drop, larger *y* -drop** | **equal *x*-drop, smaller *y* -drop** | ———— | ———— |
| $x_1$ N $x_2$ | ———— | ———— | ———— | ———— |

## 4 Two treatments, two time points, two outcome variables

Now we turn to the main theme of the present paper: how to compare the effects of two treatments, let them be denoted by A and B, both known to be effective in reducing the outcome variables x and y. The question to be investigated is whether one of the treatments has a greater effect on one of the two outcome variables, while the second treatment has a greater effect on the other.

Suppose there are *nA* patients assigned to treatment A and *nB* to B and assume that $pat_1,\ldots,pat_{nA}$ are those given treatment A and $pat_{nA+1}, \ldots, pat_{nA+nB}$ are those given B. For $1 \le i \le nA < j \le nA+nB$ consider $pat_i$ (given treatment A) and $pat_j$ (given treatment B). Depending on which of the 16 cases in the table in Section 3 that applies, we define a score $u_{ij}$ in the following way: if one of the cases

(4.1) $(x_i \text{ D } x_j \ \& \ y_i \text{ d } y_j),\ (x_i \text{ D } x_j \ \& \ y_i \text{ E } y_j),\ (x_i \text{ E } x_j \ \& \ y_i \text{ d } y_j)$

occurs we consider it as an indication that treatment A acts stronger on the outcome *x* and B stronger on *y*; in that case we define $u_{ij} = 1$. On the other hand, the cases

(4.2) $(x_i \text{ d } x_j \ \& \ y_i \text{ D } y_j),\ (x_i \text{ d } x_j \ \& \ y_i \text{ E } y_j),\ (x_i \text{ E } x_j \ \& \ y_i \text{ D } y_j)$

give the opposite indication; for each of these cases we define $u_{ij}$ = -1. For the remaining ten cases we define $u_{ij}$ = 0.

Let $T_+$ and $T_-$ denote the number of positive and negative $u_{ij}$, respectively:

$$T_+ = \sum_{i=1}^{n_A} \sum_{j=n_A+1}^{n_A+n_B} (u_{ij} = 1), \quad T_- = \sum_{i=1}^{n_A} \sum_{j=n_A+1}^{n_A+n_B} (u_{ij} = -1).$$

Clearly, if $T_+$ is considerably much larger than $T_-$ our data suggest that A reduces *x* and B reduces *y*, while the opposite information is obtained if $T_-$ is the larger quantity. Hence it is informative to consider the difference $T = T_+ - T_-$ which we can write

$$T = \sum_{i=1}^{n_A} \sum_{j=n_A+1}^{n_A+n_B} u_{ij} . \tag{4.3}$$

## 4.1 Significance testing

If, in an actual case, we have e.g. $T > 0$, the question arises how to assess the statistical significance of that result. The null hypothesis that we want to consider can be expressed in words so: if $pat_i$ belongs to the A group and $pat_j$ to the B group and if we are in one of the six cases in (4.1) and (4.2), then the probability is 1/2 that we are in (4.1), i.e.

$$H_0: P[u_{ij} = 1 \mid u_{ij} \neq 0] = 1/2.$$

A one-sided alternative can then be

$$H_1: P[u_{ij}=1 \mid u_{ij} \neq 0] > 1/2, \tag{4.4}$$

indicating that A acts stronger on the outcome *x* and B stronger on *y*; the other one-sided alternative would be

$$H_1: P[u_{ij}=1 \mid u_{ij} \neq 0] < 1/2. \tag{4.5}$$

Finally we of course have the two-sided alternative

$$H_1: P[u_{ij}=1 \mid u_{ij} \neq 0] \neq 1/2. \tag{4.6}$$

As test quantity we utilize $T$. Under the null hypothesis $T$ will be symmetrically distributed around zero, with possible values ranging from $-nA\ \ nB$ to $nA\ \ nB$; furthermore, we shall expect a rather large number of $u_{ij}$ to be zero.

In order to obtain an estimate of the null distribution of $T$ we propose the following procedure: keep all the $x$ and $y$ data but randomly permute the A and B labels among the patients; then redo the whole computation of $u_{ij}$ (where $i$ and $j$ now range over the indices of those patients who, after the random permutation, are attached to A and B, respectively) and finally compute $T$. Repeat this procedure $R$ times so that we have one authentic and $R$ pseudo-$T$ values.

If $T^{(0)}$ denotes the value of $T$ for the authentic data and $T^{(1)}$, ..., $T^{(R)}$ are the values obtained for the randomly permuted data, then an estimated p-value for a test of $H_0$ against the one-sided alternative (4.4) is

$$\frac{1}{R}\sum_{r=1}^{R}(T^{(0)} \geq T^{(r)}),$$

and of course a corresponding formula applies when testing against (4.5). Finally, an estimated p-value for testing $H_0$ against (4.6) is given by

$$\frac{1}{R}\sum_{r=1}^{R}(|T^{(0)}| \geq |T^{(r)}|).$$

**Computations related to significance testing**

In the text above the score $u_{ij}$ was defined for comparison between one patient given treatment A and one given B. It is, however, straightforward to define $u_{ij}$ in the same way also for two patients belonging to the same treatment group; the reason for doing so will be clear in a while. Obviously we will then have

$$u_{ji} = -u_{ij}$$

and in particular $u_{ii} = 0$. Thus for the intra-A sum we have

$$\sum_{i=1}^{n_A} \sum_{j=1}^{n_A} u_{ij} = 0$$

and our test quantity $T$, defined in (4.3) as

$$T = \sum_{i=1}^{n_A} \sum_{j=n_A+1}^{n_A+n_B} u_{ij}$$

can also be computed as

$$T = \sum_{i=1}^{n_A} \sum_{j=1}^{n_A+n_B} u_{ij}; \tag{4.7}$$

observe that here $j$ starts at 1.

For the estimation of the null distribution of $T$ we proposed to redo the computation of first $u_{ij}$ and then $T$ after a random permutation of the A and B labels among the patients in order to obtain pseudo-$T$ values. However, if we compute $T$ according to (4.7), it is in fact not necessary to redo all the computations, and that is why it is useful to define $u_{ij}$ for all $i$ and $j$.

To see that, introduce

$$u_{i\bullet} = \sum_{j=1}^{n_\bullet} u_{ij}$$

where $n_\bullet = nA + nB$. Then each patient has a $u_{i\bullet}$ of his or her own, and this $u_{i\bullet}$ does not depend on the distribution of A and B among the patients. Clearly $T$ is the sum of the $u_{i\bullet}$ of the patients in the A group while a pseudo-$T$ equals the sum of the $u_{i\bullet}$ corresponding to $nA$ patients drawn with simple random sampling without replacement from the $n_\bullet$ patients.

It is easy to see that the extra cost of computing $u_{ij}$ for $n_\bullet(n_\bullet - 1)/2$ pairs $(i, j)$ -- remember that the $u_{ij}$ are skew-symmetric -- is amply compensated for by our having to do it just once, not once for each permutation.

## 4.2 Interval estimation

Now we consider the problem of obtaining a measure of the difference of the effects that treatments A and B have on *x* and *y*. Among the cases where one of the treatments favours *x*, the other *y*, the proportion of cases where it is A that favours *x* and B that favours *y* is

$$\hat{\theta} = T_+ / (T_+ + T_-),$$

which is a point estimate of

$$\theta = \mathrm{P}[u_{ij} = 1 \mid u_{ij} \neq 0].$$

The distance of $\theta$ from 1/2 can clearly be used as a measure of the size of the difference between the treatments A and B.

In order to obtain a confidence interval for $\theta$ we use the bootstrap method. The main idea here is to sample independently and with replacement *nA* and *nB* observations, respectively, from the original sets of observations, that is, from the two empirical distribution functions, and thus another estimate of $\theta$, say $\hat{\theta}^*$, can be obtained. By repeating this sampling a large number, say *B*, of times, the distribution of $\hat{\theta}$ is estimated. A 1-2$\alpha$ confidence interval can be taken as the Percentile Confidence interval

$$(\hat{\theta}^*_{((B+1)\alpha)}, \hat{\theta}^*_{((B+1)(1-\alpha))})$$

where the interval endpoints are order statistics among the $\hat{\theta}^*$ values obtained during the bootstrap procedure; for technical reasons it is advantageous to choose *B* in such a way that $(B+1)\alpha$ is an integer.

Another possibility is to use what is called the Bootstrap Basic Confidence interval, which gives

$$(2\hat{\theta} - \hat{\theta}^*_{((B+1)(1-\alpha))}, 2\hat{\theta} - \hat{\theta}^*_{((B+1)\alpha)}).$$

A further possibility is to compute the basic interval after a suitable transformation of the original value and the bootstrapped ones. In this case it seems reasonable to consider $\varphi = g(\theta)$ where

$$g(\theta) = \log(\theta/(1-\theta))$$

so that the inverse is

$$\theta = g^{-1}(\varphi) = 1/(1+\exp(-\varphi)).$$

Using the notation $\hat{\varphi} = g(\hat{\theta})$ we obtain the interval

$$(g^{-1}(2\hat{\varphi} - \hat{\varphi}^*_{((B+1)(1-\alpha))}), g^{-1}(2\hat{\varphi} - \hat{\varphi}^*_{((B+1)\alpha)})).$$

## 5 Application to authentic data

In this section we apply our technique to data from a clinical trial, The Nordic Diltiazem (NORDIL) study, which was performed in order to compare a new antihypertensive agent, the calcium antagonist diltiazem, with "conventional treatment", viz. diuretics, beta-blockers, or a combination of these two. The outcome variable was occurrence of certain clinical events (Hansson *et al.*, 2000). The allocation to the two groups, "new" and "conventional", was randomized, while, in the latter group, the choice between diuretics, beta-blockers, or both, was up to the treating physician. The study had a fairly long recruitment period; the follow-up lasted between 3 and 7 years, the mean being 4.5 years. The patients were regularly checked up every six months; at such visits dose escalation as well as addition and/or change of medication could be performed according to rules laid down in the study protocol. However, our goal is not to compare how the treatments affect clinical events; instead we want to investigate whether they differ in terms of which blood pressure, systolic (SBP) or diastolic (DBP), that they are most efficient at reducing. To that end decided to use data only for those patients who started on monotherapy and did not change treatment during the period we considered. For practical reasons we restricted attention to each patient's first six months in the study. This provided us with 723 patients given diuretics, 2329 given beta-blockers, and 3711 given the calcium antagonist, assessed at baseline and after six months.

We shall consider only the comparison between the effects of diuretics and the calcium antagonist, where we get the following numbers of informative pairs:

$$T_+ = 122{,}483, \quad T_- = 40{,}916$$

where $T_+$ is the number of pairs where diuretics give a larger SBP drop and a smaller DBP drop. The test quantity is

$$T = T_+ - T_- = 81{,}567.$$

As described in Section 4.1, the data were permuted 10000 times and the pseudo test quantities calculated. None of the randomly obtained test quantities was numerically larger than the observed one and hence the two-sided p-value is less than 0.0001. There is thus a strongly statistically significant difference in how the two treatments affect the two blood pressures: diuretics reduce SBP to a greater extent while the calcium antagonist is more efficient at reducing DBP.

The effect measure $\hat{\theta}$, the proportion among informative cases where diuretics reduce SBP more and the calcium antagonist reduce DBP more, becomes 122,483/(122,483+40,916)=0.750, and the three 95 % confidence intervals for the effect measure θ are

(0.697, 0.796), (0.703, 0.802), (0.697, 0.796).

## 6 Discussion

The application of a technique for ordinal variables on variables that are manifestly quantitative, in fact on a ratio scale, may perhaps at first sight seem a bit strange; a natural objection might be that one should use the full scale properties of the variables. However, in many cases it is not obvious whether changes in such variables are most appropriately calculated as absolute or relative ones; one can easily give examples where neither of the two statements "$pat_1$ has a greater absolute reduction than $pat_2$" and "$pat_1$ has a greater relative reduction than $pat_2$" implies the other. With the method developed here there is no such dilemma.

If one decided to define “pat1 has a greater reduction than pat2” if that statement is true both absolutely and relatively, it is easy to check that this would be a less strict definition than the one presented here, i.e. our definition gives a smaller number of informative pairs. Tests performed with the both-absolute-and-relative definition and in the same manner as above would probably give results that were more statistically significant. However, the point is not to achieve higher significance but rather to make meaningful inference.

As mentioned in Section 3, comparisons of patients where one of these has a greater reduction in both outcome variables were considered uninformative for the present purpose, the reasonable explanation being that the doses of the treatments given to the two patients were not comparable. Thus, when comparing patients that are given different treatments, there will be fewer informative pairs when the doses of treatments are not adjusted to being similar in their ability to reduce the outcomes. However, the actual level of the doses given will be of no importance in our way of comparing the particular effects of the treatments. In fact, we consider it a strength of our technique that any result obtained cannot be attributed to the choice of doses given.

Furthermore, it might be feared that a huge amount of information is wasted when disregarding all non-informative comparisons. However, in the study presented here information is ignored only from very few patients. In fact, less than 1 % of the patients were not involved in any informative comparison; furthermore, about one half of these patients reacted adversely to the treatments with increases in both variables.

**Reference**

Hansson L., T Hedner, P Lund-Johansen et al.(2000). Randomized trial of effects of calcium antagonists compared with diuretics and beta-blockers on cardiovascular morbidity and mortality in hypertension: the Nordic Diltiazem (NORDIL) study. Lancet 356, 359-365.

http://journals.lub.lu.se/index.php/stat

LUND UNIVERSITY
School of Economics and Management

LUND UNIVERSITY
SCHOOL OF ECONOMICS AND MANAGEMENT
Department of Statistics
Working Papers in Statistics No 2015
Box 743
220 07 Lund, Sweden