Research Methods 2

Week 1: Document 1

Commentary on Paper by Hajek et al.

The comments on this paper will be arranged by the sections of the paper. The Abstract will not be considered, as this reiterates findings from the main paper. However, it should be remembered that in modern medical research, where finding relevant articles quickly from the large volume of research papers is of increasing importance, the abstract performs an important purpose.

Introduction

The introduction explains the background to the study. This is that mortality and morbidity after a myocardial infarction or by-pass surgery are lowered if the patient stops smoking. The authors cite articles to support this assertion but point out that the interventions studied thus far have dedicated staff and this is unlikely to be feasible within the present resources in the British National Health Service (NHS). The authors then introduced the aim of their study, which is to see if an intervention requiring limited resources, delivered in the first few days after the MI or surgery.

There are no direct statistical issues here, although it is likely that the assertions made in the cited papers will, to varying extents, rely on statistical arguments.

Method

There are five paragraphs in this section.

1. *Patient definitions and sample sizes*
2. *Patient recruitment and treatment definition*
3. *Outcome measures*
4. *Data collection*
5. *Analysis*

Patient definition and sample size.

The study ends up collecting data on more than 500 patients. However, the aim is not to learn about just these patients, it is to be able to make assertions about the relative merits of the treatments that are compared, *when applied to patients of the same 'type' as those in the study.* The idea of a *population* of patients, from which has been drawn the *sample* of patients in this study, is of fundamental importance. It is introduced in Week 4. This paragraph defines the 'type', or population of patients considered in the study.

The paragraph also indicates that the study aims to recruit two groups of patients, each of 244 patients. The rationale for this is given in terms of the probability of detecting differences between the groups. The terminology and the basis of the rationale is explained in Week 12. In broad terms, if you make the groups too small, then the natural variation that arises in small groups may be sufficient to swamp any 'real'

difference between the treatments. Therefore you should try to assess at the outset of the study how many patients you think you will need for your purposes.

Patient recruitment and treatment definition.

The design of the study is introduced here. The idea is to place patients in two groups, one receiving a 'standard' treatment (verbal advice to stop smoking plus a booklet), and the other group receiving the 'new' treatment (which comprises the standard treatment plus a more elaborate set of interventions). The patients were allocated to one or other group by random allocation. This is essentially a statistical technique, although we will not consider it in the course: further information can be found in Bland (chapter 2). In brief, random allocation is intended to create two comparable groups. At the end of the study, the effects of the treatments will be compared by comparing the groups. If the groups were formed in such a way that they were different from the start, then it would be impossible to tell whether any difference between the treatment groups observed at the end of the study was due to a difference in the treatments or to the initial difference between the groups. Random allocation is intended to overcome this problem.

This method of allocation is why this type of study is known as a randomized controlled trial.

Outcome measures

The idea of the study is to compare the effects of the treatments. This has to be done by measuring something on the patients that is relevant to the aim of the treatments. In this case several outcomes have been listed, all measuring aspects of abstinence from smoking. In this study they also asked nurses to record whether they had done everything they were suppose to do in the treatment protocols. This is because the feasibility of the intervention is also under study in this trial.

Data Collection

This largely revolves around when outcome measures were observed. The main times being at the six-week follow-up and at 12 months

Analysis

It is good practice to state at the outset of the trial how you intend to analyse the data you collect. Most of the important outcomes are *binary* variables, so c^2 *tests* were used to assess the differences between the groups. At this stage the italicised terms may well mean little to you. They will be explained in detail in Week 11. A glimpse of what is meant will also be revealed in the comments on the next section.

Logistic regression is a technique used for binary variables that is beyond the scope of this course.

Results

There are three paragraphs and two tables in this section.

1. *Baseline comparisons*
2. *Analysis of outcome*
3. *Analysis of compliance*

Baseline Comparison

This is the subject of table 1. The variables listed there describe the nature of the two treatment groups in terms of items that might well affect a patient's propensity to stop smoking, and which have been measured before the treatments were applied (i.e. at *baseline*). It is essentially a check of whether the randomization has produced two comparable groups.

Table 1 lists nine variables against which only single numbers are given. For example, there are 207 men in the control group. These nine variables are *binary* variables and are met in Week 11. Each patient can have only one of two values on these variables (man/woman, retired/not retired etc.) and the number in one of the categories is a full summary of that variable.

Four of the variables are *continuous* variables (which are met in Week 3), namely age,, no. cigarettes per day, time to first smoke in a morning, time since admission/operation. These are compared in terms of means (averages) – average age is 56 in each group. However, the degree of dispersion in the ages in each group is also measured, using the standard deviation (SD). The mean and SD are met in Week 3.

The broad conclusion is that the random allocation has produced comparable groups

Analysis of outcome

This is where the main results of the trial are revealed. The main results are in table 2, whereas the paragraph discusses some difficulties, namely how to deal with the absence of data from patients who died before follow-up or could not be contacted. There are some subtle difficulties raised here, which the authors may not have fully overcome. However, a detailed discussion is beyond our scope. It is also a useful illustration that data collected in real life can often pose problems for statistical methods.

The results in the table are given in four rows, the format of each row being identical. The first two rows relate to observations taken six weeks after admission/surgery, while the last two rows relate to data at 12 month follow-up. The first row of each pair counts those that have been continuously abstinent, which I presume means not having smoked since their MI/by-pass. The point prevalent rows relate to more limited claims of abstinence.

Each row gives the number abstaining from smoking in each group, followed in parentheses by that figure expressed as a percentage. The next figure is the difference

between the treatment groups, as a percentage and to more decimal places than the percentages in each group.

The remaining figures are items from inferential statistics, namely a 95% confidence interval (Week 7) for the difference in percentages and a P-value testing the null hypothesis that the proportions are the same in the two groups (Week 9).

The details of what these entities do will have to wait until the right week of the course. However, in fairly general terms it is this.

The abstinence rates observed are only that, the values *observed* in this study. If a different group of patients had been studied, the exact values would be a (little?) different. Does the observed difference accurately reflect the true difference that would have been obtained had a very, very large sample (effectively encompassing all patients of this type) been available? The P-value is the chance of seeing a difference as large as we have observed *if there were no 'real' difference between the groups*.

The confidence interval is easier to understand. While we have seen a difference in abstinence rates of –4.1% (last row of table), there is a 95% chance of the 'true' difference being between –12.6% and +4.5%. So while we have seen only a small difference, our data do not rule out the possibility of a difference between these limits.

Analysis of compliance

The final paragraph considers secondary matters that are not related to the difference between the treatment *per se*. The issues considered aspects of the details of the treatments (e.g. uptake of the 'buddy' system). There was also a rather sophisticated test of similarity of implementation by nurses at the different centres participating in the study. There is also a comment on whether aspects of the treatment predict its ultimate success. This is done by comparing abstinence rates amongst those who signed their commitment card (43%) versus the rate (22%) among those who did not. A hypothesis test shows that this sort of difference is unlikely to arise by chance. However, it is not clear what the results mean, as the groups are from *within* the intervention group. Therefore these groups have not been created by random allocation, so they may not be comparable at the outset, and the difference may not reflect any effect of signing the commitment card. It might, for example, that only those already determined to stop smoking are prepared to sign the card. This point is conceded in the discussion.

As the feasibility of implementing the intervention in routine clinical practice is also under examination, some descriptive information (Week 3) on the length of time needed to complete the treatment is also included (mean 34 minutes, ranging from 4 minutes to 85 minutes).

Discussion

The trial produces no clear evidence that on treatment is better than the other and the tenor of the discussion is that the intervention was not an outstanding success. This is partly related to the fact that time constraints on nursing staff prevented the full

implementation of the guidelines. Also there was an observation that such a short-term intervention might not have much chance to affect highly-dependent smokers.

The main statistical comment on the discussion is that there is a view throughout the discussion, typified by the first sentence, that the intervention was 'not effective'. However, it is never made clear why this is thought to be the case. The non-significant P-values do not reveal evidence of effectiveness but nor do they, of themselves, reveal ineffectiveness (Weeks 9 and 12). The difference in 12 month abstinence rates could be around 12% (in favour of the intervention: see 95% confidence intervals in Table 2). This may or may not represent a worthwhile improvement, but no specific comment on this is forthcoming.