

Patients, Medicines, Policies and Statistics

Simon Day

Lilly Research Centre Ltd, Windlesham, Surrey
and
Chair, PSI Public Affairs Sub-Committee.

◆INTRODUCTION◆

ON THE opening page of their book *Statistical Methods in Medical Research* (1987), Armitage and Berry acknowledge that ‘The argument is occasionally heard that statistical information contributes little or nothing to the progress of medicine, because the physician is concerned at any one time with the treatment of a single patient, and every patient differs in important respects from every other patient.’ They counter this argument with the statement: ‘the variability of disease is an argument **for** statistical information, not **against** it’ (their emphasis). When, as a student, I first read that, I was rather unimpressed. They seemed to respond with little more than the pantomime ‘Oh yes it is!’ ‘Oh no it isn’t!’ pendulum. But as my understanding of statistics grew I began to appreciate they had hit the nail square on the head. Where is statistics used in society?

- Motor car insurance: every driver drives differently but we can still make useful statements about low and high risk groups;
- Resource planning: every family has different circumstances affecting how many children they may have and the chances of survival of those children but we can make useful statements about how many school places will be needed;
- Meteorology: Who knows if it will snow on Christmas day? No-one does but the book-keepers are still prepared to accept bets on it. They know the trends and the variability from day to day: They assess the risk.

Who knows if a patient will respond to a given treatment? Nobody does. That is why we need to know whether treatments usually work, or only rarely work, or in which types of people they usually work and so on. None of the answers to these questions will give us any guarantees but at least we can kick off with some good bets rather than popping a pill at random. So the variability (or uncertainty) of disease

really is the argument **for** statistical information, and not against it.

◆DRUG DEVELOPMENT◆

The pharmaceutical industry employs statisticians in a wide variety of areas including manufacturing control, finance departments, marketing, basic laboratory research and, most commonly, clinical research. Who discovers, tests and develops new medicines? Government? Universities? Something like 98% of new medicines (the figure varies depending on which publication you read) are developed by pharmaceutical companies working in the private sector. Even so, for every 10,000 compounds that swirl around in test tubes, only one or two prove to be useful and safe enough to put onto the market so it is clearly a pretty risky business. The major areas where statisticians are employed are in the testing of new chemical entities firstly on animals, then on human, healthy, volunteers and then on ‘real’ patients. The human testing is broadly divided into four stages or ‘phases’. Phase 1 is carried out on a small number of volunteers and is primarily aimed at determining whether or not there are any major, obvious, safety problems. Phase 2 will be carried out on a small sample of patients aimed at trying to find the best dose. Phase 3 will be large scale definitive studies to demonstrate efficacy and safety. Phase 4 is often called ‘Post Marketing Surveillance’ and, as its name suggests, consists of follow-up studies of drugs after they come on to the market. In practice, the distinction between Phase 2 and 3 is often not clear. Pocock (1983) gives a good broad overview of clinical trials.

◆STOMACH ULCERS◆

.AN EXAMPLE

As an example of a study falling between Phase 2 and 3, we conducted a trial comparing four groups of patients being treated for their stomach ulcers under the following regime:

Group 1 received placebo (an inert compound)

Group 2 received active drug at dose d , twice per day

Group 3 received active drug at dose $2d$, once per day

Group 4 received active drug at dose $2d$, twice per day

So there were interesting comparisons to be made particularly between a dose of $2d$ once or twice a day and between a total daily dose $2d$ taken in one or two 'shots' per day. 500 patients were to be recruited into the study across Europe; this would give approximately 80% power to detect a difference in cure rates of about 15% between any pair of treatment groups at the 5% significance level using a simple Pearson chi-square statistic. Inevitably, at the end of the study, not all the patients could be properly assessed but Table 1 shows the simplest presentation of the efficacy results.

Table 1. Crude Cure Rates.

Treatment	Cure Rates	
Placebo	98/124	79%
d twice/day	107/126	85%
$2d$ once/day	106/130	82%
$2d$ twice/day	115/130	88%

The problem goes further than this, however. Although the results presented can be considered as valid and unbiased, they are crude. It is very likely that response rates will be different for smokers vs. nonsmokers, drinkers vs. non-drinkers, possibly men vs. women and possibly varying with age. Because the patients were assigned randomly to one of the four groups, we can be confident about the lack of bias but we should still investigate the data to try to obtain more informative estimates of cure rates.

Logistic regression is a useful way of investigating the effect of several prognostic factors on a binomial outcome. It is similar in concept to multiple regression where we model continuous outcomes. In its very simplest form, logistic regression can do the same job as a chi-squared test (i.e. the significance test) but it can also go further than that and we can build up quite complex models to describe all but the random residual variation in a process. The obvious difference between binomial proportions and continuous (normal) variates is that the former have to be restricted to the range zero to one. For this reason, instead of modelling the response rates, π , we model the function $g(\pi) = \ln\{\pi/(1-\pi)\}$ as the response variable and slot it into what looks very much like an ordinary regression model:

$$g(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 x + \beta_2 x_2 + \dots + \beta_p x_p$$

The model has to be fitted iteratively but those details need not concern us. Computer programs do that side of the work quite easily. The method allows

us to estimate the parameters (the β 's) for each of the potential risk factors (the x 's). The results we obtained are shown in Table 2.

The intercept is a bit like the intercept in an ordinary regression model and not particularly interesting. The parameter estimates themselves are not particularly interesting. They represent $\ln\{\pi/(1-\pi)\}$ which is called the log-odds ratio. Taking the exponential of these coefficients gives us the odds ratio which is interesting. We see from Table 2, for example, that the odds of being cured if you are a drinker (vs. not being a drinker) are about 1:2. (The estimated odds ratio is 1.97, 95% confidence interval from 1.3 to 3.0) So if you have an ulcer, whatever treatment the doctor prescribes, you are more likely to be cured if you are a non-drinker! 'The treatment estimates are obviously important. Compared to placebo, your odds of being cured are two and a half times greater if you are given dose $2d$ twice a day but only about one third greater if you take dose d twice a day. These estimates are, of course, not exact. If we look at the 95% confidence limits we see that even a pessimist would have to accept the value of the $2d$ twice per day regimen but the d twice per day regimen includes values less than 1 in the confidence interval so we have to accept there is not conclusive evidence of efficacy for this dose compared to placebo. The odds ratios for gender and smoking are both close to one and there is not enough evidence to conclude that these are additional risk factors. (Notice that the confidence intervals overlap unity).

The coefficient for age is rather different. For each year older you get, your odds of cure (compared to last year) drop by about 1%. Not much - but the odds of cure at age 60 vs. age 40 are about 90% (that is, $\exp(-0.005 \times 20 \text{ yrs})$). Again, perhaps this is not much. It is certainly less of an effect than that of the different treatments.

So we can conclude in order of increasing efficacy are placebo, d twice per day, $2d$ once per day, $2d$ twice per day. Whether or not you drink is an important risk factor, though your age, gender and whether or not you smoke have not been found to be risk factors.



Table 2. Results from logistic regression.

Variable	Parameter estimate	Standard Error	Odds Ratio	95% Confidence Interval for O.R.
Intercept (α)	-2.327	0.648		
Gender (β_1) Male	0			
Gender (β_1) Female	0.128	0.129	1.14	0.9 → 1.5
Smoker? (β_2) Yes	0			
Smoker? (β_2) No	-0.161	0.233	0.85	0.5 → 1.3
Drinker? (β_3) Yes	0			
Drinker? (β_3) No	0.678	0.216	1.97	1.3 → 3.0
Age (β_4)	-0.005	0.007	0.99	0.98 → 1.01
Treatment (β_5)				
Placebo	0			
d 2/day	0.300	0.270	1.35	0.8 → 2.3
2d 1/day	0.487	0.269	1.63	1.0 → 2.8
2d 2/day	0.918	0.277	2.50	1.5 → 4.3

◆THE PROCESS ◆
OF DRUG LICENSING

Having completed all our clinical trials, we are not yet able to make a new drug widely available to patients. First, and rightly so, the evidence needs to be independently assessed by government agencies who make a decision as to whether the claims being made by a pharmaceutical company are justified given the data that exists. This is a most important aspect and requires that at all stages of the clinical trials (including the analysis of the data) it is clearly stated what was done and why it was done. Many compounds never get to the market place because they are found either to be ineffective or unsafe. It is not unreasonable then for government reviewers to expect every application that is sent to them to show positive efficacy results and a good safety record. No pharmaceutical company would submit an application unless that were the case. The results presented, therefore, are not the primary evidence to consider. It is the process that produced those results that must be scrutinised. Anyone working in research in the pharmaceutical industry (statistician or not) must be not only a very good scientist but also very good at documenting their work!

◆STATISTICIANS IN THE ◆
PHARMACEUTICAL INDUSTRY

To encompass the interests and scope of work carried out by statisticians within our industry, we

have our own professional body that was formed in 1977. At the end of its first year it had 47 members. It now has approximately 600 members. We call ourselves the association of Statisticians in the Pharmaceutical Industry but with a little licence in re-arranging the initials, abbreviate this to PSI which obviously has connections with the Greek language to which we refer for much of our mathematical notation (Ψ).

Our Constitutions describes the aims of PSI as

- (a) To provide a forum for regular discussion on statistics and matters relating to the practice of statistics in the pharmaceutical industry.
- (b) To promote professional standards of statistics in matters pertinent to the pharmaceutical industry.

Each year we run scientific meetings and training courses. There are currently specialist groups working together on problems associated with laboratory data, dose proportionality, regulatory issues etc. The Public Affairs sub-committee helps with public relations and promoting the good image of statistics, statisticians and PSI (hence this article).

References

Armitage P. and Berry G. (1987) *Statistical Methods in Medical Research*, 2nd Edition Oxford: Blackwell.

Pocock S.J. (1983) *Clinical Trials: A Practical Approach*, Chichester: Wiley

Further information about PSI or working as a statistician in the pharmaceutical industry may be obtained from the author or from the PSI Executive Secretary, P0 Box 37, Ely, Cambridgeshire CB6 3XW.