PHARMACEUTICAL STATISTICS
Pharmaceut. Statist. 2002; **1**: 9–16 (DOI:10.1002/pst.002)

Changing times in pharmaceutical statistics: 1980–2000‡

MAIN PAPER

Simon Day*,†
Leo Pharmaceuticals, Simon Day, Longwick Road, Princes Risborough Buckinghamshire, HP27 9RR, U.K.

Pharmaceutical statistics has changed significantly in the last twenty years. This paper reviews some of the changes in that time, from a past dominated by the P-value to the present day. Although the last twenty years have seen significant improvements, statistical shortcomings are not a rarity, and the impact of the public perception of statistics, medicine and the pharmaceutical industry on pharmaceutical statistics is assessed. Copyright © 2002 John Wiley & Sons Ltd.

INTRODUCTION

This paper considers how the pharmaceutical industry – and statistics and statisticians within it – has changed over the last twenty years and what is likely to happen over the next twenty years. No attempt has been made to describe the earliest involvement of statistical thinking within the industry; instead, the focus is on one of the greatest highlights of the 1980s, the *P*-value.

The majority of the paper is concerned with change and the sorts of changes that may take place as natural progressions of today's changes or, indeed, the introduction of completely new ideas. The status of statistics, statisticians, medicine, physicians, pharmaceutical companies and the pharmaceutical industry are in a constant state of change. We need to capitalize on the beneficial changes and be wary of, and protective against, the detrimental ones. We need to watch our customers – and watch whatsoever they watch. Our customers look at us and make judgements. To help them make those judgements, *reality* is unimportant – *perceptions* are all important. We (the pharmaceutical industry and statisticians within it) need to look at ourselves as others see us and understand what they think of us and what they want from us.

Although this is a personal perspective driven largely by my experiences within the UK, we work in a global business and I am an avid observer of developments internationally. I welcome correspondence on the issues I raise, and those I do not raise.

A PAST OBSESSED WITH P-VALUES

In 1960, Rozeboom [1] commented that 'the stranglehold that conventional null hypothesis significance testing has clamped on publication

* Correspondence to. Simon Day, Medicines Control Agency, Room 13-205, Market Towers, 1 Nine Elms Lane, London SW8 5NQ, UK.
† E-mail: simon.day@mca.gsi.gov.uk (S. Day).
‡ This paper is an extract from a talk given at the PSI meeting on Statistics in Drug Development at the Royal Society for Arts, London, 8–9 March 2000. The second half covering the next twenty years (2000–2020) is to be published in the second issue.

Pharmaceutical STATISTICS

standards must be broken'. Forty years on things are better, but not that much better.

In a book review written in 1996, David Jones [2] made this assessment:

> 'Were I able to write my memoirs as a former member of a UK drug regulatory body, they would certainly include recollection of an image of a page in a product licence application – making a case for a new drug to be admitted to the market – consisting of a large table of *P*-values allegedly summarizing differences in treatment efficacy in a particular study, as measured by several different endpoints and within several different subgroups. Naturally, the sporadic scattering of 'significant' *P*-values amongst this pageful was difficult to interpret and did not amount to impressive evidence'.

This 'scattering' of '*P*-values' echoed an early draft of an editorial for the *British Heart Journal* by Stephen Evans who referred to '*P*-values scattered like pigeon droppings in Trafalgar Square'. Sadly, in the final version [3] his co-authors had persuaded him to take it out.

I confess that I produced many such tables of *P*-values. But often experiments were designed so poorly that with the best will in the world the statistician could not get the investigator(s) or the sponsor to pin down exactly what they were really interested in. My approach in such cases was generally to produce all the (reasonably) possible *P*-values and write a warning about multiplicity. But I confess, further, that I never really expected my warnings to be considered. The *P*-values were (and are), of course, easy to produce; even obtaining standard errors was relatively easy in those days (standard errors of treatment differences were perhaps not quite so easy). The prospect, however, of calculating and quoting confidence intervals was just too much effort and the sponsor would undoubtedly not want to pay for many hours' work to obtain them. Today I take comfort from the opinion expressed by Voss and George [4] that 'the number of tests performed ... and the level of significance should be stated to allow readers the choice of whether to [make adjustments for multiplicity]'. Certainly this was, and still is, my view as well – at least the pigeon droppings are not as common as they were. However, examples are not difficult to find. An example came to me of a study which was sponsored by and more importantly (for the purposes of the point to be made here) was analysed by statisticians from one of the largest pharmaceutical companies. It was published in 1993 and titled (and the title is important and well phrased) 'Antidepressant efficacy and cardiac safety of trimaprimine in patients with mild heart disease' [5]. Twenty-two depressed outpatients with mild heart disease entered the study. The Hamilton depression scale [6] is the only efficacy measure reported (although a clinical global impression scale is also reported as having been recorded). On days 13, 21 and 29 significant reductions ($P < 0.001$) from baseline were observed for the Hamilton depression scale. Note that there was no control group; all tests were based on within-group changes. The 22 patients are then summarized by 12 *P*-values relating to 24-hour Holter monitor data; 9 *P*-values relating to electrocardiogram data; and 20 *P*-values relating to sitting and supine, systolic and diastolic blood pressure, and heart rate (those each at four study visits). A further 20 *P*-values summarize the vital signs of the subgroup of 15 patients with coronary heart disease or angina. Table I summarizes the situation. Setting aside the efficacy data, the 22 patients were seemingly enough to demonstrate the 'safe' nature of this product based on 60 of 61 *P*-values being non-significant ($P > 0.05$).

The antidepressant paper [5] is only about eight years old and even in the relatively short time since then things have improved a lot, although still there are worrying signs. An extract from Section 2.2.2 of the ICH E9 document [7], headed 'Primary and secondary variables' reads: 'The number of secondary variables should be limited and should be related to the limited number of questions to be answered in the trial.' Figure 1 shows an extract from a protocol of a trial of 1500 patients who have suffered a stroke. It shows a clear primary objective and a few important secondary objectives. All looks in reasonable accordance with

Table I. Summary of significance tests in Cohn *et al.* [5].

Study Visits:	3	4	5	6
Efficacy (all patients)				
Hamilton depression scale		✓	✓	✓
Clinical global impression scale			(not reported)	
Safety (all patients)				
Total premature ventricular contractions	✓		✓	✓
Max premature ventricular contractions per hour	✓		✓	✓
Total premature atrial contractions	✓		✓	✓
Max premature atrial contractions per hour	✓		✓	✓
R-R interval	✓	✓		✓
QRS interval	✓	✓		✓
Q-T interval	✓	✓		✓
Sitting systolic blood pressure	✓	✓	✓	✓
Sitting diastolic blood pressure	✓	✓	✓	✓
Supine systolic blood pressure	✓	✓	✓	✓
Supine diastolic blood pressure	✓	✓	✓	✓
Heart rate	✓	✓	✓	✓
Safety (patients with coronary heart disease or angina)				
Sitting systolic blood pressure	✓	✓	✓	✓
Sitting diastolic blood pressure	✓	✓	✓	✓
Supine systolic blood pressure	✓	✓	✓	✓
Supine diastolic blood pressure	✓	✓	✓	✓
Heart rate	✓	✓	✓	✓

the ICH E9 guideline. Now, Figure 2 shows the next page of the protocol with rather more secondary objectives. Is this still a 'limited' number of secondary endpoints? Perhaps it is. But Figure 3 shows the next page of the protocol with yet more, this time tertiary, objectives. Subsequent amendments to the protocol have, not surprisingly, added to, rather than subtracted from, the number of secondary (tertiary?) endpoints.

I paint a gloomy picture, and for a more balanced picture of the past (in a much more positive light than I shed here) I recommend you look at the paper on clinical trials by John Lewis read to the Royal Statistical Society in 1983 [8]. The subtitle to the paper tells it all: 'Statistical Developments of Practical Benefit to the Pharmaceutical Industry'.

STATISTICAL SHORTCOMINGS

Regulatory statisticians have shared with those of us in the industry some of the common problems that they see in licence applications [9]; (see also [10]). Papers entitled 'Statistical shortcomings in ...' tend not to give a clear impression of the incidence of such examples, just their severity. Titles such as 'The good, the bad and the ugly' [11] from 1999 at least contain some good news, but the emphasis still seems to be with the bad (and the ugly!). Of course, now that we have more hurdles – or *guidelines* – there is greater opportunity for us to fail. But there is still plenty of room for improvement. At a voluntary inspection by the Medicines Control Agency, I was questioned on ICH E9. The inspector confided in a serious tone: 'you would be surprised how many people haven't [read E9]'. That did surprise me. The ICH E9

Pharmaceutical STATISTICS

XYZ 1234 Study	*16 December 1996*	*Page 62 of 76*
24.1	**Primary Response Criterion**	
	The percentage of patients with satisfactory functional outcome defined as score 0, 1 or 2 on modified Rankin Scale (cfr. Appendix IV) at 6 months from stroke onset.	
24.2	**Secondary Response Criteria**	
24.2.1	Median modified Rankin Scale score (cfr. Appendix IV) at 6 months from stroke onset.	
24.2.2	The percentage of patients with satisfactory functional outcome defined as score 0, 1 or 2 on modified Rankin Scale and who consider having made a complete recovery at 6 months from stroke onset.	
24.2.3	Percentage of patients with the Barthel Index > 90 points.	
24.2.4	SNSS (cfr. Appendix III) at day 10 in patients not classified as having posterior circulation infarcts (POCI, cfr. Appendix II).	
24.2.5	Median Quality of Life measurement according to SF-36 Health Survey scale at 6 months from stroke onset.	
24.2.6	Discharge destination (cfr. 18.7).	
24.2.7	Usual residency at 6 months (cfr. 18.1 & 18.7).	
24.2.8	Length of hospital stay for initial stroke.	

Figure 1. A protocol comparing two treatments for patients with acute stroke. Primary and secondary response criteria.

document is a valuable and important read. There is something in there for all of us.

PERCEPTIONS

The last twenty years have seen a shift in the public perception of statistics, medicine and the pharmaceutical industry. Let me deal with each in turn.

The Public Perception of Statistics

I think the credibility of statistics – or rather data – is not good. Recent disclosures about falsified safety records (falsified *data*) from British Nuclear Fuels' Sellafield plant [12], for example, sink data credibility to zero. On the other hand, I think that the credibility of statisticians in the public's eye has increased and will go on increasing. In the UK the credit for this lends itself in part to a better image presented by the Office for National Statistics (including its greater independence from government) or to the inability of successive governments to successfully hide statistics and data that they do not want distributed. I recall, though, the view expressed in 1996 by an agricultural statistician that 'the image of statisticians will not improve until the nexus between hypothesis

XYZ 1234 Study	*16 December 1996*	*Page 63 of 76*
24.2.9	Percentage of patients who, at 6 month follow-up answer "Yes" to the question: "Have you undertaken any paid work within the last 2 weeks?"	
24.2.10	Percentage of patients who, at 6-month follow-up answer "Yes" to the question: "Did you need any help in daily-life activities during the last 2 weeks?"	
24.2.11	Death at 6 months.	
24.2.12	Death during in-hospital treatment period (for patients withdrawn from the treatment/study: Death before day 10 from randomisation).	
24.2.13	Frequency of intracranial haemorrhage (haemorrhagic transformation or intraparenchymal haemorrhage) found on CT scan in patients with clinical deterioration or found at autopsy.	
24.2.14	Frequency of other serious bleeding complications during treatment period.	
24.2.15	Percentage of patients with substantial neurological deterioration of initial stroke (cfr. 19.3, 7).	
24.2.16	Percentage of patients experiencing adverse events.	
24.2.17	Changes in laboratory safety parameters.	

Figure 2. A protocol comparing two treatments for patients with acute stroke. Further secondary response criteria.

XYZ 1234 Study	*16 December 1996*	*Page 64 of 76*
24.3	**Tertiary Response Criteria**	
24.3.1	Causes of death.	
24.3.2	Percentage of patients with recurrent stroke (cfr. 19.2.2).	
24.3.3	Percentage of patients with deep venous thrombosis (cfr. 19.2.3).	
24.3.4	Percentage of patients with pulmonary embolism (cfr. 19.2.4).	
24.3.5	Percentage of patients with thrombocytopenia.	

Figure 3. A protocol comparing two treatments for patients with acute stroke. Tertiary response criteria.

testing and statistics is broken' [13]. Whether it be for positive reasons or negative, I am happy to jump on the bandwagon of the statistics profession getting a (slightly) better press these days. I hope this trend continues and I believe it will.

Pharmaceutical STATISTICS

The Public Perception of Medicine

The medical profession, however, has for many years been going through a difficult time in terms of public opinion. The following comments are not to be interpreted as a criticism of the medical profession (I won't jump on *that* bandwagon), but the credibility of doctors and of medicine has a substantial impact on the pharmaceutical industry and therefore on all of us who work in it.

Doctors used to be seen by the public and themselves as all powerful and all knowing. Many people (doctors and others) still hold this view, but for many it has changed. Confidence on the part of some patients to challenge the doctor has increased; perceived superiority amongst many doctors towards their patients has decreased. But it does not take much to do a lot of damage. The incidence of fraud, for example, seems to be on the increase [14], although I wonder if it really is or if there is some detection bias contributing. Our ability to detect fraud [14,15] has increased at least a little, and I think that a lot of progress will be made in this area in the future (both statistical and non-statistical). However, our *willingness* to detect and expose fraud has increased enormously [16,17] over the last five to ten years and I feel that this will continue to increase. I wish, however, that we could find a better term than 'whistle-blower': it sounds so negative and derogatory, yet these are the very people we should be thanking and praising. But lexicography is an academic luxury, given the damage often done to one's own career by exposing fraud or misconduct by others [18]. This situation has to change. Aside from fraud, there is important media coverage of, for example, an apparent excess mortality at the Bristol Royal Infirmary's cardiac unit [19]; allegations concerning retention of children's organs *post mortem* at Alder Hey Hospital in Liverpool without the permission of the coroner and without parental consent [20]; and, most notably, the conviction of Dr Harold Shipman for murdering 15 of his patients [21].

Whether or not fraud is a contributory factor in another case is as yet unknown but allegations have been made by parents that they were misled into consenting to experimental treatment to help their premature baby's breathing by the method of continuous negative extrathoracic pressure [22] at the North Staffordshire Hospital, Stoke-on-Trent. A subsequent investigation at least partly upholds these allegations [23], although I reserve judgement as to whether we have yet heard the whole truth and nothing but the truth. Whilst it may be true that participants in trials (or in this case their parents) may not easily *understand* the information they are given [24], that does not absolve the clinician from the need to explain the procedures. All of these examples impinge on public perception. Even responses prepared with the speed with which the Royal College of Pathologists [25] addressed organ retention *post mortem* do little to redress the balance. The opening sentence of that report says: 'These guidelines ... are intended ... to ensure that the examination of the body after death ... is conducted in a respectful manner *in which the public has confidence*' (emphasis added). Much of the public has already lost confidence, and certainly the publicity surrounding the original findings at Alder Hey far outweigh the publicity over the Royal College's response. These are clearly extreme cases, and nothing to do with pharmaceutical statisticians. There are less extreme cases: patients' high demand to be prescribed antibiotics for coughs and colds; doctors' over-zealous prescribing of antibiotics and the pharmaceutical companies at the end of the line churning out the antibiotics and promoting them. There are further examples, such as allegations [26] and counter-allegations [27] over an alleged link between the MMR vaccine and autism. I warn you that this is all bad news for us, the pharmaceutical industry, and each of us who work in it. When 'medicine' gets a bad press, the pharmaceutical industry is all part of it.

The Public Perception of the Pharmaceutical Industry

The pharmaceutical industry has had a very bad press in recent years. The Clinton administration did little for the pharmaceutical industry in the

USA although the industry within the UK and Europe has enjoyed better times.

I believe that the public image of the pharmaceutical industry is going down *despite* the following positive aspects:

- Quality within the industry has increased – that is the quality of the people and of the work (trials and other).
- Integrity within the industry has increased – partly through external regulatory constraints but also the attitude of the types of individuals employed.
- Academic excellence within the industry has increased – possibly only due to the 'brain-drain' from universities, although I think also for more positive reasons.

However, the *expectations* of the public have increased whilst their knowledge has not. It is not difficult to find editorials and letters in the scientific and lay press criticizing the activities of the pharmaceutical industry, and I cite particularly an editorial in the *British Medical Journal* by Dent and Hawke [28] entitled 'Too soon to market'. In summary, they are concerned that 'The unpublished data seen by the licensing authorities have not been scrutinised by the scientific community and may not have been peer reviewed which limit their ... use in licensing decisions'. The point they *may* have which would be valid is that the general scientific community has not seen the information presented to the licensing authorities. However, they appear to believe, firstly, that regulatory authorities are acting in the sponsors' interest and not that of the public; and secondly, that peer review in referred journals is an adequate process. I was pleased to see a response by Senn challenging their point of view [29] but disappointed to see four other responses essentially supporting it [30–33].

The following is taken from the preface of the book *Statistics in the Pharmaceutical Industry* [34] published over 20 years ago:

> 'We believe that the pharmaceutical industry currently documents its statistical and other evidence better than any other segment of our society. As more and more comparable problems ... become [problems] for [other] parts of industry and government ... we think that [they] will profit from learning the lessons of the pharmaceutical industry and applying them in their own spheres as appropriate'.

Clearly the gap between the knowledge and expectations of those outside the industry and the realities within the industry is widening. I do not know how to stop it continuing to do so. We must remember, though, that we are not entirely free from blame: a news headline from the *British Medical Journal* reads 'Drug company bosses jailed for selling HIV infected products' [35] and massive publicity surrounded the way clinical trial results were presented by British Biotech [36].

Little more need be said.

ACKNOWLEDGEMENTS

The thoughts presented here have been moulded over many years through many conversations with many people; I thank them all and hope those conversations and speculations will continue. Specific thanks for helpful comments regarding this paper are due to Claus Bay, Signe Birk Jensen, Professor Stephen Evans, Professor Andy Grieve, Darren Jolliffe, Professor John Lewis, David Lowson, Tony Rees, Dr James Roger, Professor John Whitehead and Dr Zoë Williams. The views expressed in this paper are not necessarily, however, the views of any of those persons. I also thank the PSI Scientific Committee for inviting me to give these views at the meeting on Statistical Issues in Drug Development and those at the meeting for their thoughtful and searching questions. They have forced me to crystallize some of my random thoughts into a presentable form.

REFERENCES

1. Rozeboom WW. The fallacy of the null hypothesis significance test. *Psychological Bulletin* 1960; **57**:416–428.
2. Jones DR. Biometrie in der Chemisch-Pharmazeutischen Industrie: Testing Principles in Clinical and Pre-Clinical Trials (book review). *Statistics in Medicine* 1996; **15**:1602–1603.
3. Evans SJW, Mills P, Dawson J. The end of the *p* value? *British Heart Journal* 1988; **60**:177–180.
4. Voss S, George S. Multiple significance tests (letter). *British Medical Journal* 1995; **310**:1073.

5. Cohn JB, Wilcox CS, Goodman LI. Antidepressant efficacy and cardiac safety of trimaprimine in patients with mild heart disease. *Clinical Therapeutics* 1993; **15**:114–126.
6. Hamilton M. A rating scale for depression. *Journal of Neurology, Neurosurgery, Psychiatry* 1960; **23**: 56–62.
7. ICH E9 Expert Working Group. Statistical principles for clinical trials: ICH harmonised tripartite guideline. *Statistics in Medicine* 1999; **18**: 1905–1942.
8. Lewis JA. Clinical trials: statistical developments of practical benefit to the pharmaceutical industry (with discussion). *Journal of the Royal Statistical Society, Series A* 1983; **146**:362–393.
9. Lewis JA, Facey KM. Statistical shortcomings in licensing applications. *Statistics in Medicine* 1998; **17**:1663–1673.
10. Jones DR. Issues in submissions to the CSM. In: Proceedings of the 12th Annual Conference, PSI: Macclesfield, 1989.
11. Facey KM. The good, the bad and the ugly—make my day! A regulatory review. In: *Proceedings of the 1999 Annual Conference*, PSI: Macclesfield, 2000.
12. Buckley C, Nuttall N, Whymant R. Sellafield chiefs face sack over safety failures. *The Times*; February 19, 2000.
13. Nester MR. An applied statistician's creed. *Applied Statistics* 1996; **45**:401–410.
14. Buyse M, George SL, Evans S, Geller NL, Ranstam J, Scherrer B, Lesaffre E, Murray G, Edler L, Hutton J, Colton T, Lachenbruch P, Verma BL. The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Statistics in Medicine* 1999; **18**:3435–3451.
15. Evans S. Statistical aspects of the detection of fraud. In: *Fraud and misconduct in medical research*, Lock S, Wells F. (eds). BMJ Publishing Group: London, 1996.
16. Lock S, Wells F (eds). *Fraud and misconduct in medical research*. BMJ Publishing Group: London, 1996.
17. Joint Consensus Conference on Misconduct in Biomedical Research. *Proceedings of the Royal College of Physicians of Edinburgh* 2000; **30** (Suppl. 7):2–3.
18. Yamey G. Protecting whistleblowers: employers should respond to the message, not shoot the messenger. *British Medical Journal* 2000; **320**:70–71.
19. Dobson R. Bristol inquiry opens. *British Medical Journal* 1999; **318**:754.
20. Smith R. Babies and consent: yet another NHS scandal (editorial). *British Medical Journal* 2000; **320**:1285–1286.
21. Dyer C. Tighter control on GPs to follow doctor's murder convictions. *British Medical Journal* 2000; **320**:331.
22. Jones J. Government sets up inquiry into ventilation trial. *British Medical Journal* 1999; **318**:553.
23. Griffiths R. *Report of a review of the research framework in North Staffordshire Hospital NHS Trust*. NHS Executive West Midlands Regional Office, 2000.
24. Elbourne D, Snowdon C, Garcia J. Subjects may not understand concept of clinical trials. *British Medical Journal* 1997; **315**:248–249.
25. Royal College of Pathologists. *Guidelines for the retention of tissues and organs at post mortem examination*. Royal College of Pathologists: London, 2000.
26. Wakefield AJ, Murch SH, Anthony A, Linnell J, Casson DM, Malik M, Berelowitz M, Dhillon AP, Thomson MA, Harvey P, Valentine A, Davies SE Walker-Smith JA. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet 1998*; **351**: 637–641.
27. Bower H. New research demolishes link between MMR vaccine and autism. *British Medical Journal* 1999; **318**:1643.
28. Dent THS, Hawke S. Too soon to market. *British Medical Journal* 1997; **315**:1248–1249.
29. Senn S. Statistical aspects of research done outside pharmaceutical industry could be improved (letter). *British Medical Journal* 1998; **316**:228.
30. Clark W, Fisher M, Misson J. Data available before products are launched are poor (letter). *British Medical Journal* 1998; **316**:228.
31. Sochart DH. Problems apply to introduction of new prosthetic implants as well as new drugs (letter). *British Medical Journal* 1998; **316**:228.
32. Williams HC. Problem is acute in dermatology (letter). *British Medical Journal* 1998; **316**:228.
33. Wolff F. Problem does not just occur with British markets (letter). *British Medical Journal* 1998; **316**:228.
34. Buncher CR, Tsay J. *Statistics in the pharmaceutical industry*. Marcel Dekker: New York, 1981.
35. Lamar J. Drug company bosses jailed for selling HIV infected products. *British Medical Journal* 2000; **320**:601.
36. Hampton JR. Clinical trial safety committees: the devil's spoon. *British Medical Journal* 2000; **320**: 244–245.

