



Editorial

Meta-analyses of the effects of dopexamine in major surgery: do all roads lead to Rome?

In this issue Gopal et al. [1] report that, in meta-analysis, dopexamine does not influence survival after major surgery or critical care. This result may seem unworthy of editorial focus – except for the fact that very recently, in another journal, Pearse et al. [2] concluded (using meta-regression analysis of substantially the same data) that low dose dopexamine dramatically halves 28-day mortality. What are the practice implications of these conflicting results? Which should we believe? Should anaesthetists routinely use dopexamine or not?

Before he became first Roman Emperor, Augustus was commissioner of roads and in 20 BC placed the *miliarium aureum* (golden milestone) near the temple of Saturn, from which all roads in the empire originated and which listed distances to the major imperial cities. From any province, however distant, it was possible to reach the Eternal City by whichever route (*omnes viae Romam ducunt* – ‘all roads lead to Rome’). Similarly in scientific enquiry given the same data, all valid methods of analysing should intuitively lead us to the same ‘answer’ of the question posed. Gopal and Pearse used different methods (or routes), which individually appeared valid, but they reached almost opposite conclusions. Some form of critical appraisal [3], including that of the relevant mathematics, is needed to help resolve the resulting dilemma.

Some technical differences in the studies

Gopal et al. [1] draw attention to some details which could plausibly have led to different conclusions from those of Pearse et al. Gopal et al. included one critical care study while Pearse et al. focussed only on operative studies. Gopal et al. found, however, that

exclusion of this data would not have altered their conclusion. There were some differences in the scoring of included trials. Pearse et al. counted 28-day mortality; Gopal et al. counted all in-hospital mortality regardless of when it occurred (which explains some of the small discrepancies in the event rates reported in the two studies).

Adding robustness to their study, Pearse et al. [2] contacted authors directly for individual patient data and thus reported both an ‘intention-to-treat’ and a ‘per-protocol analysis’. The former regards patients as being in a treatment (or control) group if they were assigned to receive the treatment (or control), regardless of whether they did so or not. The latter classifies patients into a ‘treatment’ group only if they actually received the treatment [4]. The supposed advantage of ‘intention-to-treat’ is that it avoids bias induced by patients who drop out because, say, the treatment (or the mere randomisation to it) is ineffective or causes greater harm, as might happen with side-effects which were not specifically monitored. A per-protocol approach would include these drop-outs instead in the control arm of the trial, thus missing a potentially adverse effect of (being allocated to) treatment. In fact, Pearse et al. [2] found no difference between the analyses. Gopal et al. [1] conceded they did not access original patient information, so could only use the published (i.e. per protocol) data, leaving their analysis open to the possibility of bias. However, it is impossible to predict in which direction any presumed bias may have worked.

There are several ‘decorative’ aspects of the papers that some readers may feel are important (I do not think so) [3]. Based in a well-known London teaching hospital, Pearse et al. [2] paid for professional data analysis (which presumably was of high quality) and honestly declared their interests in a company which produces/markets

dopexamine. Based elsewhere, Gopal et al. [1] declared no conflicts of interest and conducted the analysis themselves (a senior and experienced scientist being a co-author).

Pearse et al. undertook a greater number of statistical comparisons, increasing the risk of a ‘false positive’ statistical outcome: extensive sub-group analysis can yield misleadingly positive results due to chance alone [5, 6]. We now turn to some more detailed differences in analysis.

Meta-analysis vs meta-regression: relative risks vs odds ratios

A statistician will instantly recognise that differences in the papers arise in part because Gopal et al. [1] used ‘conventional meta-analysis using REVMAN 4.2 and the random effects model of DerSimonian and Laird’, while in contrast Pearse et al. [2] used ‘meta-regression where mortality was expressed as a binary outcome with a multi-level logistic approach, estimation performed using a first-level marginal quasi-likelihood model, the results of which informed a second-level predictive quasi-likelihood estimation’. The purpose of the next section is to try translate this into simpler concepts.

First, the notion of ‘heterogeneity’ is relevant [7, 8]. Where several clinical trials show broadly similar (e.g. positive) effect of treatment, it is reasonable mathematically to combine their results into a single estimate. This will have a numerical value relatively close to any one of the trials, and thus be representative of them collectively. Alternatively, the results of the trials may differ widely (e.g. one trial shows a strongly positive benefit, another equally strongly indicates great harm with treatment and so on): these trials are heterogeneous. It might be technically possible to combine the results, but numerically this ‘average’ (however computed) will have a value dissimilar from any single trial, so may be less

meaningful. An analogy can be drawn with trying to combine apples and oranges: we can certainly calculate the dimensions of the resulting hypothetical fruit, but it does not resemble anything we recognise.

It is possible to identify whether heterogeneity exists in a group of trials using several statistical tests and if it exists, there are a number of ways to deal with it:

- 1 many statisticians advise that meta-analysis should be avoided altogether or, if undertaken, any positive result viewed with great caution;
- 2 combine the results of the trials using a ‘random effects model’, but again interpret any results with caution. ‘Random effects’ refers to complex statistical methods which recognise that trials are discrete, separate and varied entities, rather than assume that they all originate from a single theoretical ‘source’ or that they are all part of an imaginary ‘mega-trial’ (which is what a ‘fixed effects’ model assumes, appropriate for a homogenous data set);
- 3 use ‘meta-regression’: this is a method that almost thrives on heterogeneity. It recognises – or indeed relies upon – true differences between trials and explores the factors or sub-groups that might contribute to these in a manner akin to logistic regression [9].

Gopal et al. [1] used a combination of (1) and (2); that is, they used a graphical (funnel) plot and a well established I^2 test of heterogeneity. They then used a random effects model but urged proper caution in the final result. Funnel plots have previously been used to show how meta-analyses can mislead, especially in small trials (e.g. the erroneous suggestion that magnesium after myocardial infarction was beneficial) [10]. In contrast, Pearse et al. did not conduct a formal test of heterogeneity (although they conceded that it did exist). Instead they used method (3), meta-regression, and found that one sub-group (low dose dopamine) yielded a significant and positive result [1].

For technical mathematical reasons meta-regression can only use data which is expressed as *odds ratios* rather

than as *relative risks*, and this is one potential source of difference between the studies. Odds ratios (as used by Pearse et al.) and relative risks (as used by Gopal et al.) are both *fractions* but of very different sorts [11].

For an event rate in the treatment group denoted p and rate in control group denoted q , relative risk (RR) is simply a ratio of the two:

$$RR = \frac{p}{q}.$$

If treatment A produces a benefit of 90% and treatment B a benefit of 10%, the relative risk is nine; i.e. A is nine times better than B.

An odds ratio (OR) is much less intuitive (which is probably why betting shops use odds rather than risk to confuse the punters). Odds are the ratio of an event occurring compared with it not occurring, or mathematically expressed as:

$$OR = \frac{p(1 - q)}{q(1 - p)}.$$

So the incidence of success vs failure with treatment A relative to the incidence of success vs failure with B is 81, a very different number from the relative risk using the same data. Odds ratios thus tend to exaggerate treatment effects, especially where the prevailing mortality or event rate is high [12]. The Cochrane Collaboration [13] prefers to avoid meta-regression when individual trials are small and heterogenous because (as with the dopexamine studies), the risk of obtaining a spurious positive result can be high. Zhang and Yu [14] used detailed mathematical argument to show specifically that great caution is needed if the odds ratio is ≤ 0.5 and the event rate is $\geq 10\%$. The data of Pearse et al. fulfil both these criteria; Gopal et al., using the more conservative measure of relative risk, avoided these problems.

The Simpson paradox

‘Decimal fractions’ (e.g. 0.125, 0.226, etc) can be manipulated just like integers so long as the decimal point is taken into account. Yet, cut a cake with a ‘decimal knife’ into three parts (of size 0.333...), put them back together and some

crumbs are always missing. By contrast, cut a cake with a ‘fractional knife’ into thirds, put it back together and we have all of our original cake. This precision of fractions and this ability to embrace geometry led ancient Pythagoreans to become fanatical in a belief that all numbers were expressible as ratios of whole numbers. However (as with odds ratios and relative risks) fractions can be dangerous. When Hippasus of Metapontum exposed that $\sqrt{2}$, although ‘real’ (e.g. the hypotenuse of a right-angled triangle whose other sides have unit length), cannot be expressed as a fraction of two integers (i.e. it is ‘irrational’), the Pythagoreans drowned him. With less dramatic consequences, Simpson [15] in 1951 extended a 1903 observation of Udny Yule [16], to show another danger: where

$$\frac{A}{B} > \frac{a}{b} \text{ and } \frac{C}{D} > \frac{c}{d},$$

it does not necessarily follow that

$$\frac{(A + C)}{(B + D)} > \frac{(a + c)}{(b + d)}.$$

So if we are presented with a series of fractions such as mortality data in a series of clinical trials where the ‘event rate’ is the numerator (e.g. A , a , C and c) and the ‘total number of patients’ the denominator (e.g. B , b , D and d), we might obtain a perverse result if we simply combine the numerators ($A + C$ and $a + c$) and the denominators ($B + D$ and $b + d$), especially if the denominator data are heterogenous. A simple example is as follows [17]: in two trials treatment A produces benefit in 93% of patients (81/87) in the first trial and 73% (192/263) in the second trial. Treatment B produces benefit in just 87% (234/270) and just 69% (55/80) respectively. Clearly A outperforms B? Not necessarily since the total benefit from A is 273/350 (78%) while B is 289/350 (83%).

Formal meta-analytical techniques supposedly avoid Simpson’s paradox (e.g. trial selection, proper weighting, random effects models, etc) but no simple or specific ‘test’ assesses whether the paradox has confounded the data analysis [18]. Statistical discussion of it can be

elaborate [17] with even experienced researchers disagreeing [19–22]. One consensus is that, where different ways of combining the data produce different results, the paradox may be responsible. It is then important to look at the data in another (usually graphical) way, focusing especially on possible covariates or ‘confounding influences’ [23].

A graphical approach to the data: L’Abbé plot

To lend overview to increasingly complex meta-analytical data, L’Abbé et al. suggested a simple graphical plot of the data and several versions of their original suggestions have been developed [24, 25]. For the cardinal result of each of our two studies, we can plot the mortality with drug treatment against the

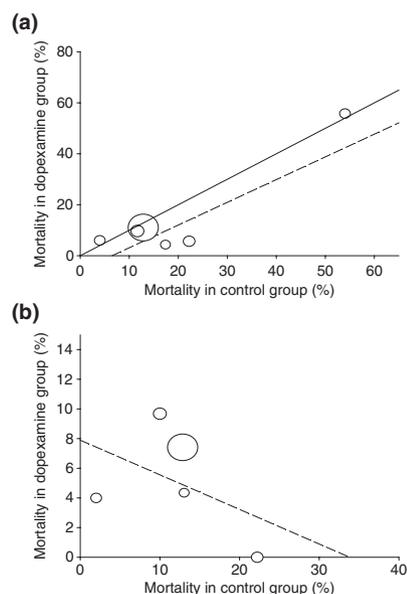


Figure 1 (a) Data from the study of Gopal et al. Mortality in the dopexamine groups (%) plotted against mortality in control groups (%), with the size of each trial proportional to the size of the symbol. The solid line indicates identity (equal mortality between dopexamine and control groups). The dashed line is the linear regression line of best fit though the data points. (b) Data from the study of Pearse et al. for low dose dopexamine (their intention-to-treat data are used, but the analysis is unchanged using their per-protocol data). For clarity no line of identity is shown. The dashed line is the linear regression line of best fit though the data points.

mortality in the control group (the latter being the ‘prevailing mortality’, a higher mortality indicating a higher-risk cohort). If drug treatment has no effect, then we expect the mortality with drug to be identical to control. For the data used by Gopal et al. (Fig. 1a), dopexamine consistently reduces mortality by approximately 3–5%, regardless of the prevailing mortality. It is plausible that a drug might have this modest effect. In contrast, the data used by Pearse et al. yields a curiosity (Fig. 1b). Dopexamine appears remarkably beneficial when prevailing mortality is very high, but paradoxically worsens mortality in a healthy cohort (Fig. 1b). This is difficult to explain and creates an internal inconsistency for the Pearse paper that does not exist within the Gopal paper. Critics who suggest that there are too few data points for me to draw any regression lines in Fig. 1 are reminded that I have simply plotted the same data as used by the authors to substantiate their main claims.

Conclusions

Both papers have their merits and any combination of factors mentioned above may have conspired to yield the contradictory results. Without repeating both analyses from scratch, it is impossible to say which factor is most responsible. However, Fig. 1 is persuasive: available data do not support a large beneficial effect of dopexamine. Feynman [26] cautioned that ‘extraordinary claims need extraordinary evidence’ and the claim that dopexamine reduces mortality by 50% is certainly extraordinary, placing this drug above aspirin/streptokinase post-myocardial infarction in terms of magnitude of treatment effect on a disease process [27]. The evidence underpinning this claim however falls short of persuasive.

This is not to say that dopexamine is useless. Both Gopal et al. and Pearse et al. – experts in the care of high risk surgical patients – feel that dopexamine is worthy of investigation and both suggest examining low doses. Probability theory originated with gambling [28] so it is appropriate to wager that a comprehensive trial will find that, where prevailing mortality is approximately 15–20% in a high risk cohort, dopexamine in optimum dose at best

reduces this by approximately 5% (Fig. 1a). It is impossible to say if this will justify its widespread use, and many other issues such as side effects will need consideration.

Readers should have their own views but the discussion above demonstrates how, when faced with conflicting advice, critical appraisal can help us arrive at what is at least a reasoned opinion [3]. From whichever road one starts, it is possible to reach Rome. All one needs is the right map.

Acknowledgements

I thank Professor Henry McQuay, Nuffield Professor of Clinical Anaesthetics, University of Oxford, for his helpful comments on the manuscript.

J. J. Pandit

Consultant Anaesthetist,
John Radcliffe Hospital,
Oxford, UK

E-mail: jaideep.pandit@dpag.ox.ac.uk

References

- Gopal S, Jayakumar D, Nelson PN. Meta-analysis on the effect of dopexamine on in-hospital mortality. *Anaesthesia* 2009; **64**: 589–94.
- Pearse RM, Belsey JD, Cole JN, Bennett ED. Effect of dopexamine infusion on mortality following major surgery: individual patient data meta-regression analysis of published clinical trials. *Critical Care Medicine* 2008; **36**: 1323–9.
- Pandit JJ, Yentis SM. All that glitters... How to assess the ‘value’ of a scientific paper. *Anaesthesia* 2005; **60**: 373–83.
- Lachin JM. Statistical considerations in the intent-to-treat principle. *Controlled Clinical Trials* 2000; **21**: 167–89.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society (Series B: Methodological)* 1995; **57**: 289–300.
- Counsell CE, Clarke MJ, Slattery J, Sandercock PAG. The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis? *British Medical Journal* 1994; **309**: 1677–81.
- Song F, Sheldon TA, Sutton AJ, Abrams KR, Jones DR. Methods for

- exploring heterogeneity in meta-analysis. *Evaluation and the Health Professions* 2001; **24**: 126–51.
- 8 Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analysis. *British Medical Journal* 2003; **327**: 557–60.
 - 9 Morton SC, Adams JL, Suttorp MJ, Shekelle PG. *Meta-regression Approaches: What, Why, When, and How?* (Technical Review 8), Rockville, MD: Agency for Healthcare Research and Quality, 2004.
 - 10 Egger M, Davey Smith G. Misleading meta-analysis: lessons from “an effective, safe, simple” intervention that wasn’t. *British Medical Journal* 1995; **310**: 752–4.
 - 11 Bland JM, Altman DG. The odds ratio. *British Medical Journal* 2000; **320**: 1468.
 - 12 Davies HTO, Kinloch I, Tavakoli M. When can odds ratios mislead? *British Medical Journal* 1998; **316**: 989–91.
 - 13 Diversity and Heterogeneity. The Cochrane Collaboration. <http://www.cochrane-net.org/openlearning> (accessed 6 December 2008).
 - 14 Zhang J, Yu KF. What’s the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes *Journal of the American Medical Association* 1998; **280**: 1690–1.
 - 15 Simpson EH. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society Series B* 1951; **13**: 238–41.
 - 16 Yule GU. Notes on the theory of association of attributes in statistics. *Biometrika* 1903; **2**: 121–34.
 - 17 Reintjes R, de Boer A, van Pelt W, Mintjes-de Groot J. Simpson’s paradox: an example from hospital epidemiology. *Epidemiology* 2000; **11**: 81–3.
 - 18 Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. *Journal of Clinical Epidemiology* 1995; **48**: 71–9.
 - 19 Hanley JA, Thériault G. Simpson’s paradox in meta-analysis. *Epidemiology* 2000; **11**: 613.
 - 20 Cates C. Simpson’s paradox and calculation of number needed to treat from meta-analysis. *BMC Medical Research Methodology* 2002; **2**: 1.
 - 21 Moore RA, Gavaghan DJ, Edwards JE, Wiffen P, McQuay HJ. Pooling data for number needed to treat: no problems for apples. *BMC Medical Research Methodology* 2002; **2**: 2.
 - 22 Altman DG, Deeks JJ. Meta-analysis, Simpson’s paradox, and the number needed to treat. *BMC Medical Research Methodology* 2002; **2**: 3.
 - 23 Schield M. Three little graphs: three graphs to promote statistical literacy. In: M Niss, ed. *Proceedings of the 10th International Congress on Mathematical Education*. Denmark: Roskilde University, 2004, pp. 1–7.
 - 24 L’Abbe K, Detsky A, O’Rourke K. Meta-analysis in clinical research. *Annals of Internal Medicine* 1987; **107**: 224–33.
 - 25 Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. *British Medical Journal* 1996; **313**: 735–8.
 - 26 Feynman RP. *Six Easy Pieces*. New York: Addison-Wesley, 1995.
 - 27 Baigent C, Collins R, Appleby P, Parish S, Sleight P, Peto R. ISIS-2: 10 year survival among patients with suspected acute myocardial infarction in randomised comparison of intravenous streptokinase, oral aspirin, both, or neither. *British Medical Journal* 1998; **316**: 1337–43.
 - 28 Pandit JJ. Editorial: gambling with ethics? A statistical note on the poisson (binomial) distribution *Anaesthesia* 2008; **63**: 1163–6.

Copyright of *Anaesthesia* is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.