

In response to: Rolfe G. (2008) In response to Paley J., Cheyne H., Dalglish L., Duncan E. & Niven C. (2007) Nursing's ways of knowing and dual process theories of cognition. *Journal of Advanced Nursing* 62(2), 268–269.

The null hypothesis: a reply

Professor Rolfe, having denied that he and Freshwater made the mistake we accused them of (Paley *et al.* 2007), promptly makes it again (Rolfe 2008). Towards the end of his JAN FORUM piece, he says: 'evidence-based treatments are predicated on the *failure* to find evidence that confirms the null hypothesis' (p. 268). In other words, he thinks that evidence-based treatments are justified by a *lack of evidence* for the null hypothesis, rather than by *evidence against* the null hypothesis. This is precisely the belief we attributed to him. Moreover, when he quotes a passage from Freshwater & Rolfe (2004) designed to show that we have misinterpreted him and his co-author, he omits the previous sentence: 'Postpositivist scientific evidence, then, is purely negative; it consists in...showing that there is no evidence to support the null hypothesis'. Please note: 'showing there is *no evidence to support* the null hypothesis', not 'providing *evidence against* the null hypothesis'. Far from 'falsely representing' the argument of the book (Rolfe 2008, p. 269), we merely reported what Freshwater and Rolfe actually said, and what Rolfe has now explicitly repeated. It is surreal. Rolfe complains that he and Freshwater have been misrepresented, but insists (in order to justify this accusation) that they said precisely what we said they said.

The obvious question is: why does Rolfe continue to assert that 'scientific evidence consists in showing that there is no evidence to support the null hypothesis' even as he denies that he *has* asserted it? His JAN FORUM piece suggests an answer to this question. He appears to think that experimental design in the clinical and psychological sciences is based on the ideas of Karl Popper. He does not say so in as many words, but the implication is unmistakable, and his lecture on Popper would be pointless otherwise. However, this view is mistaken. Experimental design is in fact based on the work of Fisher, Neyman and Pearson (Fisher 1935, Neyman & Pearson 1966); and although the Fisher and Neyman–Pearson approaches are distinct, they are complementary, and routinely combined in the standard textbooks (Lehmann 1993). Fisher's *p* values are, of course, familiar, while

null hypothesis testing is frequently referred to as Neyman–Pearson testing (Mayo & Spanos 2006).

It is true that Popper, Fisher, Neyman and Pearson all share a broadly falsificationist outlook. However, there is a major difference between Popper's views and those of the other three. According to Popper, if a hypothesis, *H*, survives an attempt to falsify it, the most that can be said is: it survived. The fact of *H*'s survival in this one test tells us nothing about its future performance; it tells us only about its past performance. We cannot infer its truth, and we are not warranted in relying on it. Any suggestion to the contrary just smuggles induction back in; and, as Rolfe correctly observes, Popper rejects the idea of induction *tout court*. Moreover, further attempts to falsify *H* do not alter the situation. If *H* survives additional tests, then it survives. But we are still not entitled to accept it, or to behave as if it were true.

In contrast, the frequentist statistical models of Fisher, Neyman and Pearson are designed to permit inductive inference. For them, it is not enough to say that *H* survived; they want to identify conditions under which it is possible to claim that *H* survived, *and* that it is unlikely that *H* would have survived if it were false. If these conditions are met, according to frequentist theory, then an inductive inference is licensed. In which case, it is legitimate to 'accept' *H* (and in this context, 'accept' is a technical term referring to what Neyman calls 'inductive behaviour'). This acceptance of *H* is subject to the results of further experiments; but these further experiments are, crucially, part of what *accepting H* licenses.

Neyman–Pearson tests achieve these conditions by adopting a Binomial model of experiment in which a null hypothesis, H_0 , is tested against an alternative, H_1 . The key feature of H_0 is that it specifies what would occur purely by chance. In a randomized controlled trial, for example, H_0 is the hypothesis that there will be no difference between the arms of the trial, which (assuming a large enough sample size) is what would happen by chance – that is to say, what would happen if the treatment being evaluated were ineffective. The statistical procedures adopted in an analysis

of the experimental data determine whether the results of the trial are consistent with H_0 or not. This is a matter of probability: because H_0 represents chance, the procedures can determine how likely it is that the experimental results would have been obtained if H_0 were true. If the results are inconsistent with H_0 (in probability terms: they are unlikely to have occurred by chance), the null hypothesis is 'rejected', and H_1 , the hypothesis that the treatment under consideration is effective, can be 'accepted'. Among the various forms of inductive behaviour which, in this case, 'reject H_0 ' licenses, further experimental evaluations of the treatment will be particularly significant.

On the evidence of both the book and his JAN FORUM piece, Rolfe does not really understand what a null hypothesis is. He appears to believe that it is simply the converse, the mere negation, of a primary hypothesis which the Popperian strategy attempts to disprove (his example of a null hypothesis in the book is 'What goes up does not come down'). But it is not. It is a specification of what would happen by chance; and a properly conducted experiment, if it finds evidence against the null hypothesis, can license the rejection of the latter, and consequently legitimate the (always provisional) conclusion that the treatment works. To suppose, as Professor Rolfe does, that a clinical trial is an attempt to discover 'evidence for the null hypothesis', with the null hypothesis being construed purely as the converse of a 'positive' alternative, is to make a fundamental error about the statistical models used in experimental design.

John Paley

Senior Lecturer

Department of Nursing & Midwifery, University of Stirling,
Stirling, UK

e-mail: jhp1@stir.ac.uk

Helen Cheyne

Research Programme Leader

Nursing, Midwifery and Allied Health Professions
Research Unit, University of Stirling, Stirling, UK

Len Dalglish

Professor of Decision Making

Department and Nursing and Midwifery, University of Stirling,
Stirling, UK

Edward Duncan

Clinical Research Fellow

Nursing, Midwifery and Allied Health Professions
Research Unit, University of Stirling, Stirling, UK

Catherine Niven

Director

Nursing, Midwifery and Allied Health Professions
Research Unit, University of Stirling, Stirling, UK

References

- Fisher R.A. (1935) *Design of Experiments*. Oliver and Boyd, Edinburgh.
- Freshwater D. & Rolfe G. (2004) *Deconstructing Evidence-Based Practice*. Routledge, London.
- Lehmann E.L. (1993) The Fisher, Neyman–Pearson theories of testing hypotheses: one theory or two? *Journal of the American Statistical Association* 88, 1242–1249.
- Lehmann E.L. (1995) Neyman's statistical philosophy. *Probability and Mathematical Statistics* 15, 29–36.
- Mayo D.G. & Spanos A. (2006) Severe testing as a basic concept in a Neyman–Pearson philosophy of induction. *British Journal for the Philosophy of Science* 57, 323–357.
- Neyman J. & Pearson E.S. (1966) *Joint Statistical Papers*. University of California Press, Berkeley, CA.
- Paley J., Cheyne H., Dalglish L., Duncan E. & Niven C. (2007) Nursing's ways of knowing and dual process theories of cognition. *Journal of Advanced Nursing* 60, 692–701.
- Rolfe G. (2008) In response to Paley J, Cheyne H, Dalglish L, Duncan E & Niven C (2007) Nursing's ways of knowing and dual process theories of cognition. *Journal of Advanced Nursing* 62(2), 268–269.