RUNNING HEAD: Views about the best evidence

Do the Public Share Practitioners' Views about the Best Evidence?

Elizabeth J. Robinson[1], Deborah Biggerstaff[2], Sally Jennings[1] and Elizabeth A. Maylor[1]

[1] Department of Psychology, University of Warwick, UK.

[2] Institute of Clinical Education, University of Warwick, UK.

Corresponding author: Professor E.J. Robinson, Dept of Psychology, University of Warwick, Coventry CV4 7AL, UK.

Email: e.j.robinson@warwick.ac.uk

Tel: +44/0 24 765 23096; Fax: +44/0 24 765 24225

Patient Education and Counseling, in press, March 2012.

ACKNOWLEDGEMENTS

Sally Jennings was able to contribute to this research thanks to the support of an undergraduate bursary from the Experimental Psychology Society awarded to Elizabeth Robinson.

We are grateful to Nicola Doherty for help with recruiting older adults.

ABSTRACT

Objective. To investigate whether general practitioners (GPs) and patients agree on what constitutes the best evidence for the effectiveness of treatments.

Method. GPs and members of the public aged 18-83 read five scenarios describing comparisons between hypothetical treatments for common ailments. Each scenario reported that one treatment was the more effective, as determined by randomised controlled trial (RCT), audit of treatment outcomes from many doctors' patients, a single doctor's clinical experience, a friend's experience, or a web-based sales site. Participants rated how confident they would be that the treatment reported to be more effective would work for them.

Results. All participants had least confidence in the web-based sales site, more confidence in a friend's experience and more still in one doctor's experience. For doctor's experience, audit and RCT, amongst the public there were some differences by age but, importantly, only GPs had most confidence in evidence from an RCT.

Conclusion. GPs may treat evidence from RCTs as the gold standard while members of the public (their patients) may not afford it that same respect.

Practice Implications. GPs engaged in shared decision-making should be alert to possible differences from their patients in the weight given to different types of evidence.

Do the Public Share Practitioners' Views about the Best Evidence?

1. Introduction

In clinical policy and practice in the UK and US, it is currently considered important that decisions about treatments should be (i) based on the best available evidence and (ii) shared between patient and practitioner. The research reported here examines the relationship between these two aspects of current best practice by asking the question: Do practitioners and members of the public share the same views about what constitutes reliable evidence? With advice and information about medical conditions widely available in the media, it may not be uncommon for patients to have formed views about their preferred treatment before they consult their doctor. Suppose patients place trust in sources of evidence that their doctors do not consider reliable. Shared decision-making is likely to be impeded, especially if the mismatch in degree of trust remains implicit during the consultation. As a first step in identifying the extent of any such problem, we asked general practitioners (primary care physicians) and members of the public (potential patients) to indicate their trust in five different sources of evidence concerning the effectiveness of hypothetical treatments for a variety of common ailments. The sources of evidence were: randomised controlled trial (RCT), retrospective study of outcomes of many doctors' treatment decisions (audit), a single doctor's clinical experience, a friend's experience, and the recommendation from a web-based sales company.

RCTs are considered by many clinicians, researchers, funding bodies and policy makers to represent the gold standard for evidence based medicine, although some have urged that other forms of evidence not be under-valued [1;2;3;4]. Evidence from retrospective studies of non-randomised allocation to treatments by a large number of practitioners (here we use the term audit), is widely considered to be less compelling than that

from RCTs, although some researchers dispute claims of bias [2;3;4]. Bias to a much greater extent can arise in the clinical experience of a single practitioner. Hence it is relatively uncontroversial to rank these three sources of evidence about the effectiveness of treatments in the order: (i) RCT, (ii) audit, and (iii) a single doctor's experience. Even less controversial is to rank a friend's experience as a less reliable source, while a source with a commercial interest such as a web-based sales site might reasonably be judged less reliable still.

Our expectation was that both doctors and members of the public would doubt the reliability of a web-based sales site, and we included this in the study reported below as a check that participants treated the task as intended. What was not obvious, however, was how members of the public would evaluate the other four sources of evidence. In particular, would they consider the evidence from an RCT to be more trustworthy than evidence from an audit (both of which were described without the use of these labels), or than the experience of a single doctor, or of a friend with experience of the same ailment?

There are grounds for predicting that members of the public might not accord evidence from an RCT particular standing. Many studies suggest that participants in RCTs struggle to understand that allocation to treatment arms is at random [5;6;7;8;9]. Even when trial participants do understand about their own random allocation, they may provide an incorrect rationale such as scarcity of resources [6;8]. If they understood and accepted the scientific rationale for randomisation, we would not expect the very deep-seated difficulties consistently reported in the literature. Results consistent with this suggestion come from studies involving healthy participants (potential patients) who responded to written research scenarios [10]. In this research, participants judged that researchers would find out just as much about the relative effectiveness of two treatments if treatment allocation were based on doctors' and patients' joint preferences, rather than at random. Participants failed to understand the scientific benefits of random allocation as opposed to allocation on the basis

of doctors' and patients' preferences, despite best efforts to explain the rationale to them (e.g., the explanation for random allocation then recommended by UK's Multicentre Research Ethics Committees). Importantly, the problem is not due simply to failure to understand what random allocation means: participants accurately identified random and non-random methods of allocation to two groups [11]. Rather, participants struggled to grasp the rationale for making comparisons between groups of people as opposed to what works for each individual, and providing explanations did not reduce their difficulty.

People's difficulties accepting the rationale for evaluating treatment effectiveness using currently preferred research designs may impact on treatment decisions in standard consultations. This possibility has not been explored hitherto, but an implication is that patients arrive at their consultation hoping to be offered treatments that may not be supported by clinical evidence that meets the practitioner's standards. When discussing treatment options with their practitioner, there may be a clash of perspectives about which each party could remain unaware, and which interferes with the process of shared decision-making about treatments.

Shared decision-making is justified on the grounds that patients are more likely to be satisfied with their consultation and more likely to comply with agreed treatment regimes when they have played an active role in discussion about their condition and about treatment options [12;13;14]. Despite this, patients' preference for a shared or a direct approach has been found to vary with the medical condition under discussion, as well as with personal characteristics such as social class and age [15;16]. The UK Government's White Paper (2010) [17] promised to make shared decision-making 'the norm' on the grounds that this improves patients' satisfaction, increases adherence to the chosen treatment and reduces costs (p. 13, p. 45). Despite the value placed on shared decision-making by policy-makers, doctors face challenges in taking into account patients' preferences [18]. It has been argued that a

necessary condition "for evidence-based patient choice is that the patient accurately comprehends a clear synthesis of high-quality, balanced empirical reports about the clinical diagnosis, the available therapeutic options, and their inherent advantages and disadvantages – that is the 'objective evidence'" (p. 123) [13]. Yet hitherto there has been no investigation, so far as we are aware, of the extent to which patients and practitioners agree on the relative reliability of different types of evidence for treatment effectiveness.

Of course we cannot assume that there is uniformity of view amongst patients on the reliability of evidence from different sources. As a check on the extent of variation amongst members of the public (potential patients), we included participants from 18 to 83 years of age and with a wide range of educational and occupational backgrounds. We predicted that the older participants would show relatively high trust in the view of an experienced doctor, because they could have been accustomed to a paternalistic style that was more common in years gone by [19;20;21]. Interestingly, this same prediction about trust in sources of evidence arises from the finding that older people prefer to rely on experts' opinions rather than take an active part in decision-making [22], and in particular that they tend to prefer a direct rather than a shared approach to medical decision-making [16].

In summary, we predicted that GPs would give highest confidence ratings to a treatment that had proved more successful in an RCT, with declining confidence for the other forms of evidence in the order: audit based on many doctors' treatment decisions, a single doctor's experience, a friend's experience, and finally, web-based sales recommendation. We predicted that members of the public would differentiate less clearly than GPs between these forms of evidence, and in particular that the older adults would have relatively high confidence in a doctor's experience.

2. Method

2.1 Ethics. Ethical approval for the research was granted by Warwick University's Social Science and Humanities Research Ethics Committee. The research was conducted in accordance with the British Psychological Society's ethical guidelines. Participants gave informed written consent to participate.

2.2 Participants ***and Recruitment.*** General practitioners (GPs, N = 47) were recruited by virtue of their studying courses at a UK University Medical School. At the end of suitable classes, GPs were invited to stay on to complete the questionnaire. No record was kept of the precise number who were unable or unwilling to stay, but this was well under 20%. Of those who agreed to participate, 44 revealed their ages: M_{age} = 47.6 years, *SD* = 7.9. Thirty-two revealed their gender: 16 female and 16 male.

Members of the public (potential patients), none of whom had occupations within the health services, were (i) a sample of 56 first-year psychology undergraduates who gained course credit for taking part, and who were allocated at random to take part in various ongoing research studies in the psychology department. No-one declined. Those included were 47 females and 6 males, 3 gender not revealed, M_{age} = 19.3 years, *SD* = 0.65, 2 age not revealed; (ii) a convenience sample of 50 mature adults aged 23 to 63 years, recruited personally by the third author from the local neighbourhood and workplaces. One person declined. Those included were 25 females and 25 males, M_{age} = 41.3 years, *SD* = 11.6; and (iii) a sample of 61 older adults aged at least 64, recruited from a UK Psychology Department's panel of volunteers for research into ageing. One person declined and one failed to complete the questionnaire. Those included were 39 females and 19 males, 3 gender not revealed, M_{age} = 71.9 years, *SD* = 5.5.

The GPs had practised for a mean of 17.2 years, *SD* = 7.8. The undergraduates all had 18+ school leaving qualifications; 82% of them had at least one science or mathematics subject. The mature adults had a range of qualifications: 24% had their highest qualification

at age 16, 18% at 18, 34% at degree level and 24% at postgraduate level. The corresponding percentages for the older adults were 30%, 15%, 34%, and 21%, respectively.

2.3 *Materials and Procedure.* Participants were invited to complete a questionnaire that asked them to imagine that they were suffering with ailments: headache, cough, indigestion, eczema and back pain. They read five hypothetical scenarios, each reporting the relative effectiveness of two treatments (A or B, C or D, P or Q, S or T, or X or Y) from different forms of evidence: RCT, audit, doctor's clinical experience, friend's experience or web-based sales recommendation. The order of the five scenarios, the labels for the treatments and which of the pair was more effective were counterbalanced between participants. See Table 1 for the wording for each form of evidence. The description of the RCT was closely based on reports of randomised controlled trials in the UK quality press (*Guardian*; *Times*; *Daily Telegraph*). As is typical in such reports, there was no quantitative information about the absolute or relative effectiveness of the two treatments under comparison. The descriptions of the other forms of evidence were matched for style and approximate number of words. Within these constraints, we aimed to describe each evidence type accurately, succinctly and in such a manner that the key differentiating features between the types would be grasped readily by participants. With these aims in mind we did not provide precise details of the subject populations, but used phrases such as 'indigestion like yours' implying relevance to the participant's imagined complaint (see Table 1). Participants indicated how confident they would be that the treatment deemed more effective would work for them, using a scale from 1 (not at all confident) to 5 (very confident), with no wording on the intermediate numerical points.

Table 1 near here

All participants completed the questionnaire individually, on paper, typically taking no more than 10 minutes. The GPs, undergraduates and the older adults were in small groups,

sitting well separated and in the presence of the Experimenter. The mature adults were either alone with the Experimenter (N = 14) or completed the questionnaire in their own time and returned it to the Experimenter a few days later, confirming they had not discussed their answers with anyone else (N = 36).

3. Results

Table 2 shows the mean confidence scores shown by each participant group for each of the five types of evidence. Our strategy for testing the predictions was as follows. First, we conducted overall analyses to discover whether, as predicted, there was a significant interaction between participant group and evidence type, with the GPs differing from the other groups, particularly with respect to RCTs. We then examined each of the four participant groups separately to discover whether confidence scores differed for the five evidence types. Finally, we examined the possible influences of additional factors, namely, gender, GPs' experience, and educational qualifications.

Table 2 near here.

Participants' confidence scores were analysed using a 4x5 mixed-design ANOVA with a between-subjects factor of participant group (GPs, undergraduates, mature adults, older adults) and a within-subjects factor of evidence type (RCT, audit, doctor's experience, friend's experience, web-based sales recommendation). Here and elsewhere where there were departures from sphericity, Greenhouse-Geisser corrections to degrees of freedom and *p* values are reported. Main effects of participant group, $F(3, 210) = 8.34$, $MSE = 1.50$, $p < .001$, $\eta p^2 = .11$, and evidence type, $F(3.3, 688.0) = 160.06$, $MSE = 0.76$, $p < .001$, $\eta p^2 = .43$, were, as predicted, qualified by an interaction between participant group and evidence type, $F(9.8, 688.0) = 3.16$, $MSE = 0.76$, $p < .001$, $\eta p^2 = .04$. To locate the source of the interaction, first, we conducted a sequence of four 3 x 5 (participant group x evidence type) ANOVAs, each omitting one of the participant groups. The interaction between participant group and

evidence type appeared only when GPs were included, suggesting that the GPs differed from the other participant groups in their pattern of confidence in the different types of evidence. Second, a sequence of five 4 x 4 ANOVAs, each omitting one of the evidence types, showed that the interaction appeared only when RCT was included, suggesting that the participant groups differed in their confidence in RCT relative to the other types of evidence. Taken together, these findings suggest that it is the GPs' trust in RCTs that is responsible for the original two-way interaction.

For each of the four participant groups, we conducted paired samples *t*-tests (significance level: $p < .05$) to test whether it showed descending order of trust in the evidence types: RCT, audit, doctor's experience, friend's experience and web-based sales recommendation. Our prediction was that the GPs would show this pattern of decreasing trust, but that the other three participant groups would differentiate less clearly between evidence types. Since we were testing against a predicted order, there was no need to correct for making the four planned comparisons. GPs showed relative trust that was closely in line with the predictions, although the audit evidence from many doctors was treated with no more confidence than the single doctor's experience: RCT > audit; audit = doctor's experience; doctor's experience > friend's experience; and friend's experience > web-based sales recommendation: $t(46) = 2.47$, $p = .017$; $t(46) = 0.59$, $p = .585$; $t(46) = 5.73$, $p < .001$; and $t(46) = 3.96$, $p < .001$, respectively. Also broadly in line with the predictions, the three groups of members of the public differentiated less clearly between evidence types. For undergraduates, paired comparisons showed: RCT = audit; audit = doctor's experience; doctor's experience > friend's experience; and friend's experience > web-based sales recommendation: $t(55) = -1.15$, $p = .253$; $t(55) = -1.31$, $p = .196$; $t(55) = 6.92$, $p < .001$; and $t(55) = 4.77$, $p < .001$, respectively. For mature members of the public: RCT < audit; audit = doctor's experience; doctor's experience > friend's experience; friend's experience > web-

based sales recommendation: $t(49) = -2.91$, $p = .005$; $t(49) = 1.22$, $p = .227$; $t(49) = 3.24$, $p = .002$; and $t(49) = 7.41$, $p < .001$, respectively. For the older adults: RCT = audit; audit < doctor's experience; doctor's experience > friend's experience; friend's experience > web-based sales recommendation: $t(60) = 1.00$, $p = .321$; $t(60) = -3.09$, $p = .003$; $t(60) = 6.52$, $p < .001$; and $t(60) = 5.19$, $p < .001$, respectively.

Supplementary analyses. (i) There were no significant effects of gender. (ii) There were no significant effects of GPs' years of experience. (iii) Concerning educational qualifications, the difference between the GPs' and members of the public's patterns of trust in the different sources of evidence was not due merely to GPs having a generally higher level of educational qualifications: we selected mature and older adults with postgraduate qualifications (N = 25) and compared their confidence scores against the GPs'. Independent samples *t*- tests revealed no significant differences for any of the evidence types (all *t*s < 1) except for RCT. For RCT, the GPs showed significantly higher trust than the members of the public with postgraduate qualifications, $t(37.29) = 2.76$, $p < .01$.

4. Discussion and Conclusions

4.1 Discussion

As expected, members of the public differed from the GPs in their relative confidence in the effectiveness of treatments evaluated by different sources of evidence, as shown by the interaction between participant group and evidence type. Furthermore, it appears to be the GPs who were distinctively different from the three groups of members of the public, as shown by the absence of interaction when the GPs were excluded from the analysis. Moreover, the GPs' distinct pattern was specific to RCTs, as shown by the absence of interaction when RCT was excluded from the analysis. Also as expected, all four participant groups showed the lowest confidence in the recommendation from the web-based sales site, which we included as a check that participants understood the task as intended.

Reassuringly, all participant groups had more confidence in a doctor's experience than in a friend's experience. They also all had more confidence in the latter than in the web-based sales. The differences between GPs and members of the public lay in the relative weight given to RCT, audit and doctor's experience. As predicted, GPs' confidence was highest in evidence from the RCT, with lower confidence in the audit which, contrary to prediction, was given no more weight than a single doctor's experience, although the mean confidence ratings are in the expected direction. All the GPs in our sample were studying courses in a University Medical School, and some were GP trainers, so they may have been atypically well versed in current views of a 'hierarchy of trust' that puts evidence from RCTs at the top [23].

In contrast to the GPs, none of the groups of members of the public gave the highest confidence ratings to the RCT. As predicted, the older adults had more confidence in the single doctor's experience than the audit of treatment outcomes from many doctors' patients, which in turn was rated no differently from the RCT. The mature adults gave lower confidence ratings to the RCT than to the audit and the doctor's experience, which did not differ from each other. The undergraduates showed no difference in their confidence ratings to the RCT, audit and doctor's experience. Overall, then, there is no sign that members of the public judged RCTs to provide the gold standard for evidence about the effectiveness of treatments.

This result is consistent with the findings arising from trial participants and from healthy adults' judgments about hypothetical trials: as mentioned in the introduction the scientific rationale for randomisation seems not to be readily understood, and many adults judge that doctors would find out just as much about the effectiveness of treatments from non-randomised designs [10]. Hitherto, such findings have been seen as important for ensuring that recruits into trials give adequately informed consent, for optimising recruitment,

and for understanding why some trial participants are upset and angry when they subsequently discover that they or their family members were randomised [8]. The research reported here highlights an implication that has so far been ignored, namely, the possible impact of failures to understand the scientific rationale for randomising on decision-making in standard consultations that have no research component. Unlike GPs, members of the public (potential patients) appear to place no particularly high value on evidence from RCTs as compared with clinical audit and a single doctor's experience, and so might prefer a treatment whose effectiveness has not been demonstrated in a way that their GP finds convincing.

In addition to any difficulty understanding the scientific advantages of random allocation, the older adults in our sample showed particular trust in the experience of a doctor who, according to the scenario 'has built up an impression of how well her patients do on each of the treatments.' This trust in the doctor's view is unlikely to lead to disagreements between patient and doctor about proposed treatment, as even if the doctor's view were based on an RCT (in which the older adults had relatively low confidence), the patient can defer to the doctor's view. The more important implication is for shared decision-making, in which older patients may be unwilling to engage [16]. The policy of making shared decision-making the norm might not be welcomed by older patients in particular.

4.2 Conclusions

Prior to this study, it was unknown to what extent doctors and members of the public agreed or disagreed about what constitutes the best evidence for the effectiveness of treatments. It could have been, for example, that members of the public had greater confidence in a friend's experience than in more objective evidence. This was not the case. To that extent, the findings give no cause for alarm that doctors and patients have fundamentally different views on what constitutes good evidence for the effectiveness of

treatments. Yet as far as evidence from RCTs is concerned, doctors need to be alert: while they may treat RCTs as the gold standard, their patients may not.

4.3 Practice Implications

When decision-making about treatments includes discussion about the evidence for the effectiveness of treatments, doctors might check on whether they and their patient agree on their trust in such evidence, and the relative weight they are willing to attach to it. Older patients in particular might place more weight on the doctor's personal experience than on RCTs, while the doctor might place more weight on evidence from RCTs than on a doctor's personal experience.

REFERENCES

[1] Conceto, J., & Horwitz, R.I. (2004). Beyond randomised versus observational studies. *Lancet, 363,* 1660-1661.

[2] Dixon-Woods, M. (2007). Appraising qualitative research for inclusion in systematic reviews. *Journal of Health Services Research and Policy, 13,* 56.

[3] Kaptchuk, T.M., (2001). The double-blind, randomized, placebo-controlled trial: Gold standard or golden calf? *Journal of Clinical Epidemiology, 54,* 541-549.

[4] Worrall, J. (2002). What is evidence-based medicine? *Philosophy of Science, 69, S316-S330.*

[5] Appelbaum, P.S., Grisso, T., Frank, E., O'Donnell, R.N., & Kupfer, D. J. (1999). Competence of depressed patients for consent to research. *American Journal of Psychiatry, 156,* 1380-1384.

[6] Featherstone, K., & Donovan, J. L. (2002). "Why don't they just tell me straight, why allocate it?" The struggle to make sense of participating in a randomised controlled trial. *Social Science & Medicine, 55,* 709-719.

[7] Lidz, C.W., Appelbaum, P.S., Grisso, T., & Renaud, M. (2004). Therapeutic misconception and the appreciation of risks in clinical trials. *Social Science & Medicine, 58,* 1689-1697.

[8] Snowdon, C., Garcia, J., & Elbourne, D. (1997). Making sense of randomisation; responses of parents of critically ill babies to random allocation of treatment in a clinical trial. *Social Science & Medicine, 45,* 1337-1335.

[9] Wade, J., Donovan, J.L., Lane, J.A., Neal., D.E., & Hamdy., F.C. (2009). It's not just what you say, it's also how you say it: Opening the black box of informed consent

appointments in randomised controlled trials. *Social Science &Medicine, 68,* 2018-2028.

[10] Robinson, E. J., Kerr, C., Stevens, A., Lilford, R., Braunholtz, D., & Edwards, S. (2004). Lay conceptions of the ethical and scientific justifications for random allocation in clinical trials. *Social Science & Medicine, 58,* 811-824.

[11] Kerr, C., Robinson, E.J., Stevens, A., Braunholtz, D., Edwards, S., & Lilford, R. (2004). Randomisation in trials: do potential trial participants understand it and find it acceptable? *Journal of Medical Ethics, 30,* 80-84.

[12] Coulter, A., & Ellins, J. (2007). Effectiveness of strategies for informing, educating and involving patients. *British Medical Journal, 335*, 24-27.

[13] Llewellyn-Thomas, H.A. (2009). Values clarification. In A. Edwards & G. Elwyn (Eds.), *Shared Decision-Making in Health Care* (pp 123-133). Oxford: OUP.

[14] Sackett, D.L., Haynes, R.B., Gibson, E.S., Hackett, B.C., Taylor, D.W. et al., (1975). Randomised clinical trial of strategies for improving medication compliance in primary hypertension. *Lancet, 305*, 1025-7.

[15] Frosch, D.L., & Kaplan, R.M., (1999). Shared decision making in clinical medicine: Past research and future directions. *American Journal of Preventive Medicine, 17*, 285-294.

[16] McKinstry, B. (2000). Do patients wish to be involved in decision making in the consultation? A cross sectional survey with video vignettes. *British Medical Journal, 321***,** 867–871.

[17] Secretary of State for Health. (2010). *Equity and Excellence: Liberating the NHS.* UK Stationary Office.

[18] Say, R.E., & Thomson, R. (2003). The importance of patient preferences in treatment decisions – challenges for doctors. *British Medical Journal, 327*, 542-545.

[19] Beisecker, A. (1988). Aging and the desire for information in medical decisions: patient consumerism encounters. *Gerontologist, 28*, 330–355.

[20] Coulter, A. (2002). *The* a*utonomous patient: Ending paternalism in medical care*. London: Nuffield Trust.

[21] Secretary of State for Health. (2000). *The NHS plan: A plan for investment, a plan for reform*. London: The Stationery Office.

[22] Mather, M. (2006). A review of decision making processes: Weighing the risks and benefits of aging. In L. L. Carstensen & C. R. Hartel (Eds.), *When I'm 64: Committee on aging frontiers in social psychology, personality, and adult developmental psychology* (pp. 145-173). Washington, DC: National Academies Press.

[23] Greenhalgh, T. (2001). *How to read a paper: The basics of evidence-based medicine*. London, UK: BMJ Books.